1994

Pattern Recognition In The Structure Of Strings Of Characters Using Multivariate Statistical Analysis

Muhammad Syamsun

PATTERN RECOGNITION IN THE STRUCTURE OF STRINGS OF CHARACTERS USING MULTIVARIATE STATISTICAL ANALYSIS

by

Muhammad Syamsun

Department of Computer Science

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Faculty of Graduate Studies
The University of Western Ontario
London, Ontario
March 1994

ISBN 0-315-93190-6

ABSTRACT

Pattern recognition approach is used in this research as an alternative to univariate and ruled-based systems for recovering source text encrypted in monoalphabetic and polyalphabetic substitution ciphers. The approach uses a multivariate statistical technique called discriminant analysis to classify encrypted characters as source text symbols. It requires the generation of invariant measurement vectors for all groups of characters from the source and the encrypted text to classify and to identify characters. The measurement vectors consist of the values of variables generated from one-graph and digraph structures of the text.

The research suggests that quadratic discriminant analysis, which requires more time and space in the process of computation is not better than linear discriminant analysis. Furthermore, the stepwise linear discriminant procedure is useful in selecting variables in the measurement vector and in reducing the number of redundant variables. The use of digraph structures of text improves the total and individual percentages of correctness for the groups of characters in comparison to the use only of one-graph structures. They also improve the results of recovering the plain or source text.

The research suggests that the use of trigraph structures will further improve the percentages of correctness and should be taken into consideration for continuing this research.

ACKNOWLEDGEMENTS

I wish to sincerely thank my supervisor Dr. John M. Carroll, for his constant help, guidance, support, understanding, and encouragement given to me throughout the course of this work. Without his vast knowledge of the subject and willingness to share his ideas and valuable suggestions, this piece of work would not have been possible.

As well, to my wife, Sri Andarini, and my children, Wiji Sundari, Indra Wijiadi, Tri Indradi, and Fitri Meilani, I dedicate this thesis for their love and moral support given to me from miles away.

I would like also thank the Second Development University Project, the Ministry of Education and Culture, Republic of Indonesia, for providing the scholarship over the years of my studies at the University of Western Ontario. This period of time constituted a significant part of my personal and intellectual development.

CONTENTS

CERTIFICATE OF EXAMINATION ii

ABSTRACT iii

ACKNOWLEDGEMENTS iv

CONTENTS v

LIST OF ILLUSTRATIONS viii

LIST OF TABLES x

CHAPTER 1 INTRODUCTION 1

1.1 Pattern and Variation in The Process of Identification 1

1.2 Model of Pattern Identification 3

1.3 A Simple Example 5

1.3.1 The components involved in the discriminant analysis 6

1.3.2 Example of discriminant calculator using only one variable (relative frequency of character) 6

1.4 Identification of Letters in English Texts 13

1.4.1 Monoalphabetic substitution problem 13

1.4.2 Polyalphabetic substitution problem 17

1.5 Research objectives 19

CHAPTER 2 STRUCTURES OF STRINGS OF CHARACTERS 22

2.1 Introduction 22

2.2 Some Measures from 1-graph and 2-graph Structures 29

2.2.1 Statistical measures 29

2.2.2 Information theoretic measures . . . 31
2.3 Relationship between source text and crypto text . . . 34
2.3.1 Monoalphabetic substitution . . . 34
2.3.2 Polyalphabetic Substitution . . . 52

CHAPTER 3 DISCRIMINANT ANALYSIS AND CLASSIFICATION 58
3.1 Introduction . . . 58
3.2 General Ideas . . . 60
3.3 Fisher's Method For Two Populations . . . 62
3.3.1 Fisher's linear discriminant function . . . 62
3.3.2 Fisher's sample linear discriminant function . . . 64
3.3.3 Hypothesis testing . . . 66
3.4 The General Problem of Classification . . . 67
3.4.1 Bayes Procedure for Two Populations . . . 67
3.5 Classification with Two Multivariate Normal Populations . . . 71
3.5.1 The case of $\Sigma_1 = \Sigma_2 = \Sigma$ . . . 71
3.5.2 The case of $\Sigma_1 \neq \Sigma_2$ . . . 74
3.6 Evaluating Classification Functions . . . 76
3.6.1 Known population parameters . . . 76
3.6.2 Unknown population parameters . . . 76
3.6.3 General case . . . 77
3.6.4 Partition of Total Sample and Jackknifing . . . 78

CHAPTER 4 RESEARCH METHODS AND MATERIALS 80
4.1 Introduction . . . 80
4.2 Bayes Procedure for Several Populations . . . 81
4.2.1 The Minimum Expected Cost of Misclassification Method . . . 81
4.2.2 Classification with Normal Populations . . . 83
4.2.3 General Evaluation . . . 88
4.2.4 Implementation consideration for Bayes Procedure . . . 89
4.3 Fisher's Method for Several Populations : Canonical Analysis . . . 89
4.3.1 Using Fisher's Discriminants to Classify . . . 93
4.4 Stepwise Discriminant Analysis . . . 94

4.4.1 Analysis for multiple populations 93
4.4.2 The stepwise algorithm 94
4.4.3 Implementation consideration for Fisher's Method and Stepwise Procedure 97
4.5 Implementation 97

CHAPTER 5 COMPUTATIONS AND RESULTS **103**
5.1 Introduction 103
5.2 Variable Selection in Stepwise Linear Discriminant 107
5.3 Comparison of Results 109
5.4 Examples of Recovering Encrypted Texts 110

CHAPTER 6 CONCLUSIONS AND SUGGESTION **128**
6.1 Methodological point of view 128
6.2 Prescriptive point of view 129
6.3 Suggestive point of view 130

APPENDIX A : Results from Linear Discriminant Analysis **132**

APPENDIX B : Results from Quadratic Discriminant Analysis **208**

APPENDIX C : Results from Multivariate Cryptanalysis **257**

REFERENCES **282**

VITA **285**

LIST OF ILLUSTRATIONS

1.1 Measurement of character pattern 3

1.2 Diagram of pattern identification 4

1.3 Diagram of feature selection 4

1.4 Sample distribution of relative frequency for some characters 12

1.5 A sample of source text 14

1.6 Crypto text from the sample text after monoalphabetic substitution 15

1.7 Crypto text from the sample text after polyalphabetic substitution 18

2.1 Structural Signature of Text 39

2.2 Sample distribution of 12 variables of 5 characters involved in linear discriminant analysis 40

4.1 The example of BMDP7M commands for stepwise linear discriminant analysis of monoalphabetic substitution by using twelve variables 100

4.2 The example of BMDP7M commands for stepwise linear discriminant analysis of polyalphabetic substitution, position 3 by using twenty variables 102

5.1 Source text, example 1 118

5.2 Source text, example 2 119

5.3 Crypto text, example 1, monoalphabetic substitution 120

5.4 Crypto text, example 1, polyalphabetic substitution 121

5.5 Crypto text, example 2, monoalphabetic substitution 122

5.6 Crypto text, example 2, polyalphabetic substitution 123

5.7 Cryptanalysis of monoalphabetic substitution using linear discriminant, 12 variables, example 2 124

5.8 Cryptanalysis of monoalphabetic substitution by using quadratic discriminant, 12 variables, example 2 125

5.9 Cryptanalysis of polyalphabetic substitution using linear discriminant, 20 variables, example 2 126

5.10 Cryptanalysis of polyalphabetic substitution using quadratic discriminant, 20 variables, example 2 127

C.1 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 1 variable, Example 1 258

C.2 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 3 variables, Example 1 259

C.3 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 12 variables, Example 1 260

C.4 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 1 variable, Example 1 261

C.5 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 3 variables, Example 1 262

C.6 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 12 variables, Example 1 263

C.7 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 1 variable, Example 1 264

C.8 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 3 variables, Example 1 265

C.9 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 20 variables, Example 1 266

C.10 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 1 variable, Example 1 267

C.11 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 3 variables, Example 1 268

C.12 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 20 variables, Example 1 269

C.13 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 1 variable, Example 2 270

C.14 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 3 variables, Example 2 271

C.15 Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 12 variables, Example 2 272

C.16 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 1 variable, Example 2 273

C.17 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 3 variables, Example 2 274

C.18 Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 12 variables, Example 2 275

C.19 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 1 variable, Example 2 276

C.20 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 3 variables, Example 2 277

C.21 Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 20 variables, Example 2 278

C.22 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 1 variable, Example 2 279

C.23 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 3 variables, Example 2 280

C.24 Cryptanalysis of Polyalphabetic Substitution Using Quadratic Discriminant, 20 variables, Example 2 281

List of Tables

1.1 Sample average of relative frequencies 5

1.2 Coefficients of linear discriminant for character groups by using one variable . 7

1.3 Classification matrix for the training sample using only one variable 8

1.4 Classification matrix for the validating sample using only one variable 10

1.5 One of the rules of monoalphabetic substitution 13

1.6 Information for each group of letters 16

1.7 One of the rules of polyalphabetic substitution with key length = 3 19

1.8 Information for each letter type 20

2.1 Frequency and Relative Frequency of Characters from Different Strings Based on 1-graph Structure 22

2.2 Information available when a particular character is followed by any other character in 2-graph structure 24

2.3 Information available when a particular character follows any other character in 2 graph structure 25

2.4 Two-dimensional Array of Frequency for String-1 26

2.5 Two-dimensional Array of Frequency for String-2 26

2.6 Two-dimensional Array of Relative Frequency for String-1 27

2.7 Two-dimensional Array of Relative Frequency for String-2 27

2.8 Two-dimensional Probability Space from 2-graph Structure 27

2.9 Two-dimensional Probability Space from a 2-graph Structure (modified) . . . 28

2.10 Row-wise Conditional Probability Space from 2-graph Structure 30

2.11 Column-wise Conditional Probability Space from 2-graph Structure 31

2.12 Row-wise Analysis of Variance from Two-dimensional Probability Space . . . 32

2.13 Column-wise Analysis of Variance from Two-dimensional Probability Space . 33

2.14 One-graph table of frequency and relative frequency of characters in the source text 34

2.15 One-graph table of frequency and relative frequency of characters in the crypto text 35

2.16 Digraph frequency generated from source text and crypto text 36

2.17 Digraph relative frequency generated from source text 37

2.18 Digraph relative frequency generated from crypto text 37

2.19 Digraph of frequency representing self-dependency for position 1 in source text and crypto text 54

2.20 Digraph of frequency representing self-dependency for position 2 in source text and crypto text 54

2.21 Digraph of frequency representing self-dependency for position 3 in source text and crypto text 55

2.22 Digraph of frequency representing dependency between positions 1 and 2 for source text and crypto text 55

2.23 Digraph of frequency representing dependency between positions 2 and 3 for source text and crypto text 56

2.24 Digraph of frequency representing dependency between positions 3 and 1 for source text and crypto text 56

4.1 Prior Probability Estimate of Every Character in Source Text 98

5.1 Coefficient of linear discriminant for character groups, monoalphabetic, 1 variable 104

5.2 Classification matrix for the training sample : monoalphabetic 1 variable . . . 105

5.3 Classification matrix for the validating sample : monoalphabetic 1 variable . 107

5.4 Summary of Tables 5.2 and 5.3 for character groups, monoalphabetic, 1 variable 112

5.5 Variable used in linear discriminant analysis, 3 variables 112

5.6 Variable used in linear discriminant analysis, 12 variables 113

5.7 Variable used in linear discriminant analysis, 20 variables 113

5.8 Summary of correctness for the validating sample : monoalphabetic substitution 114

5.9 Summary of correctness for the validating sample : polyalphabetic substitution, position 1 115

5.10 Summary of correctness for the validating sample : polyalphabetic substitution, position 2 116

5.11 Summary of correctness for the validating sample : polyalphabetic substitution, position 3 117

A.1 Coefficients of linear discrimination for character groups, monoalphabetic, 1 variable 133

A.2 Classification matrix for training sample : monoalphabetic 1 variable 134

A.3 Classification matrix for validating sample : monoalphabetic 1 variable 136

A.4 Coefficients of linear discriminant functions for character groups : monoalphabetic, 3 variables 138

A.5 Classification matrix for training sample : monoalphabetic 3 variables 139

A.6 Classification matrix for validating sample : monoalphabetic 3 variables . . . 141

A.7 Coefficients of linear discriminant functions for character groups : monoalphabetic, 12 variables 143

A.8 Classification matrix for training sample : monoalphabetic 12 variables 147

A.9 Classification matrix for validating sample : monoalphabetic 12 variables . . 149

A.10 Coefficients of Linear Discrimination for Character Groups, Polyalphabetic Position 1, 1 variable 151

A.11 Classification matrix for training sample : Polyalphabetic Position 1, 1 variable 152

A.12 Classification matrix for validating sample : polyalphabetic position 1, 1 variable 154

A.13 Coefficients of Linear Discriminant Functions for Character Groups : Polyalphabetic, Position 1, 3 Variables 156

A.14 Classification matrix for training sample : polyalphabetic position 1, 3 variables 157

A.15 Classification matrix for validating sample : polyalphabetic position 1, 3 variables 159

A.16 Coeficients of linear discriminant functions for character groups : polyalphabetic, position 1, 20 variables 161

A.17 Classification matrix for training sample : polyalphabetic position 1, 20 variables 166

A.18 Classification matrix for validating sample : polyalphabetic position 1, 20 variables 168

A.19 Coefficients of linear discrimination for character groups, polyalphabetic, position 2, 1 variable 170

A.20 Classification matrix for training sample : polyalphabetic position 2, 1 variable 171

A.21 Classification matrix for validating sample : polyalphabetic position 2, 1 variable 173

A.22 Coefficients of linear discriminant functions for character groups : polyalphabetic, position 2, 3 variables 175

A.23 Classification matrix for training sample : polyalphabetic position 2, 3 variables 176

A.24 Classification matrix for validating sample : polyalphabetic position 2, 3 variables 178

A.25 Coefficients of linear discriminant functions for character groups : polyalphabetic, position 2, 20 variables 180

A.26 Classification matrix for training sample : polyalphabetic position 2, 20 variables . . . 185

A.27 Classification matrix for validating sample : polyalphabetic position 2, 20 variables . . . 187

A.28 Coefficients of linear discrimination for character groups, polyalphabetic, position 3, 1 variable . . . 189

A.29 Classification matrix for training sample : polyalphabetic position 3, 1 variable 190

A.30 Classification matrix for validating sample : polyalphabetic position 3, 1 variable . . . 192

A.31 Coefficients of linear discriminant functions for character groups : polyalphabetic, position 3, 3 Variables . . . 194

A.32 Classification matrix for training sample : polyalphabetic position 3, 3 variables 195

A.33 Classification matrix for validating sample : polyalphabetic position 3, 3 variables . . . 197

A.34 Coeficients of Linear Discriminant Functions for Character Groups : Polyalphabetic, Position 3, 20 variables . . . 199

A.35 Classification matrix for training sample : polyalphabetic position 3, 20 variables . . . 204

A.36 Classification matrix for validating sample : polyalphabetic position 3, 20 variables . . . 206

B.1 Classification matrix for training sample : quadratic discriminant, monoalphabetic, 1 variable . . . 209

B.2 Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 1 variable . . . 211

B.3 Classification matrix for training sample : quadratic discriminant, monoalphabetic, 3 variables . . . 213

B.4 Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 3 variables . . . 215

B.5 Classification matrix for training sample : quadratic discriminant, monoalphabetic, 12 variables . . . 217

B.6 Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 12 variables . . . 219

B.7 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 1 variable . . . 221

B.8 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 1 variable . . . 223

B.9 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 3 variables . . . 225

B.10 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 3 variables . . . 227

B.11 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 20 variables 229

B.12 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 20 variables 231

B.13 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 1 variable 233

B.14 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 1 variable 235

B.15 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 3 variables 237

B.16 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 3 variables 239

B.17 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 20 variables 241

B.18 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 20 variables 243

B.19 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 1 variable 245

B.20 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 1 variable 247

B.21 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 3 variables 249

B.22 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 3 variables 251

B.23 Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 20 variables 253

B.24 Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 20 variables 255

CHAPTER 1

INTRODUCTION

1.1 Pattern and Variation in The Process of Identification

Variation is the basis for the discrimination and classification of groups of objects. Without variation, one would not be able to identify objects. Variation in groups of objects can be divided into two parts : (a) *between-groups variation*, and (b) *within-group variation*.

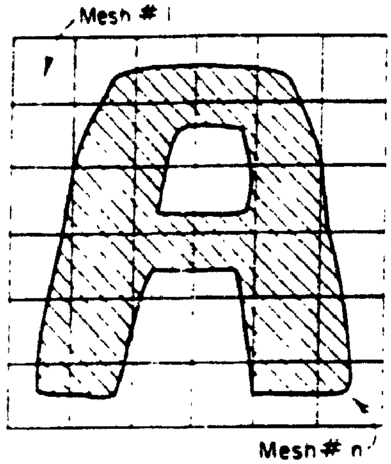
In order to discriminate and classify groups of objects it is necessary to have certain threshold value of the ratio of between-groups variation to within-group variation. This principle was introduced by Fisher [1936] to discriminate and classify some groups or varieties of iris plants by observing the pattern of flowers of the plants.

Before discriminating and classifying groups of objects, one usually investigates certain patterns of the objects that can give variation. In other words, one should try to recognize certain patterns from the objects through one or more observable *features* . In iris plants, Fisher (1936) found that it was possible to discriminate and classify by using the features of the petal and sepal of the iris flowers. Furthermore, if the observable features are measurable, then it will be possible to discriminate and classify quantitatively. Fisher [1936] then measured four *random variables* which are considered as the representation of the pattern of iris flowers. The variables are : (a) petal length ($\mathbf{x}_1$) , (b) petal width ($\mathbf{x}_2$), (c) sepal length ($\mathbf{x}_3$), and (d) sepal width ($\mathbf{x}_4$). These variables form a *feature vector* or *measurement vector* denoted by $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4)$. He measured these variables by taking fifty samples from each of three groups or varieties of iris : (1) *iris setosa*, (2) *iris versicolor*, and (3) *iris virginica*. He then generated some functions in the form of linear combinations of the original variables. These functions were obtained by maximizing the ratio of between-groups and within-group variations, and they were called *Fisher's multivariate linear discriminant functions*. The functions can then be used to identify an unknown iris plant as one of the three varieties. If one has measurement vector data of an iris flower he can identify, with a certain level of confidence, the group or variety to which it belongs. The measurement vectors taken from the samples of varieties which were used to build the functions can also be used to test the *discriminating power* of the functions. But, usually it is more highly recommended to use some other samples from known groups or varieties to test the goodness of the functions.

If a feature cannot give enough variation to discriminate and classify groups of objects, then it will be necessary to obtain some other possible features from the objects. The sum of variation generated by all of the features is called *total variation*. It is expected that the total variation will give discriminating power to identify the objects uniquely. In the case of iris plants, the four variables gave enough information or variation to do the classification. If they do not provide enough variation, then one needs to obtain other measurable features, such as leaf length, plant height, and so on.

Another example in relation to variation is the pattern recognition of letters or characters [Fukunaga, 1972]. Groups of letter A's can be identified through the measurement of grey-level of some numbered meshes that constitute letter A. Every mesh becomes a random variable with grey-level as the measurement value (see Figure 1.1). The feature or measurement vector $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ consists of grey-level values for n random variables. $\mathbf{X}$ is also called the *random vector*. The groups of letters B, C, and so on, have the same number of random variables with certain grey-level values. Multivariate discriminant functions can be built by taking samples from every group of letters, and these functions can then be used to identify a measurement vector for one of the groups of letters.

In brief, the aim of pattern recognition is to automate processes performed by humans [McLachlan, 1992].



$$\Longrightarrow \quad X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x(1) \\ x(2) \\ \vdots \\ x(n) \end{bmatrix}$$

Figure 1.1: Measurement of character pattern

1.2 Model of Pattern Identification

Identifying an object as belonging to one of several groups of objects through the measurement vector of the object can be represented by the diagram in Figure 1.2 [Duda *et al*, 1973].

The main problems in identifying an object based on the model diagram are :

(a) to determine the random variables in the measurement vector which can give enough variation to enable one to distinguish the groups of objects.

(b) to build the discriminant calculators and the maximum selector.

(c) to make the decision to classify the object into one of the groups.

It is not unusual to involve so many variables in the measurement vector that the process of identification becomes very complicated and difficult. To overcome this situation it is necessary to transform the n original variables (in n-dimensional space) into a lower m-dimensional space ($m \ll n$). This transformation is called the *feature selection process* [Fukunaga, 1972].

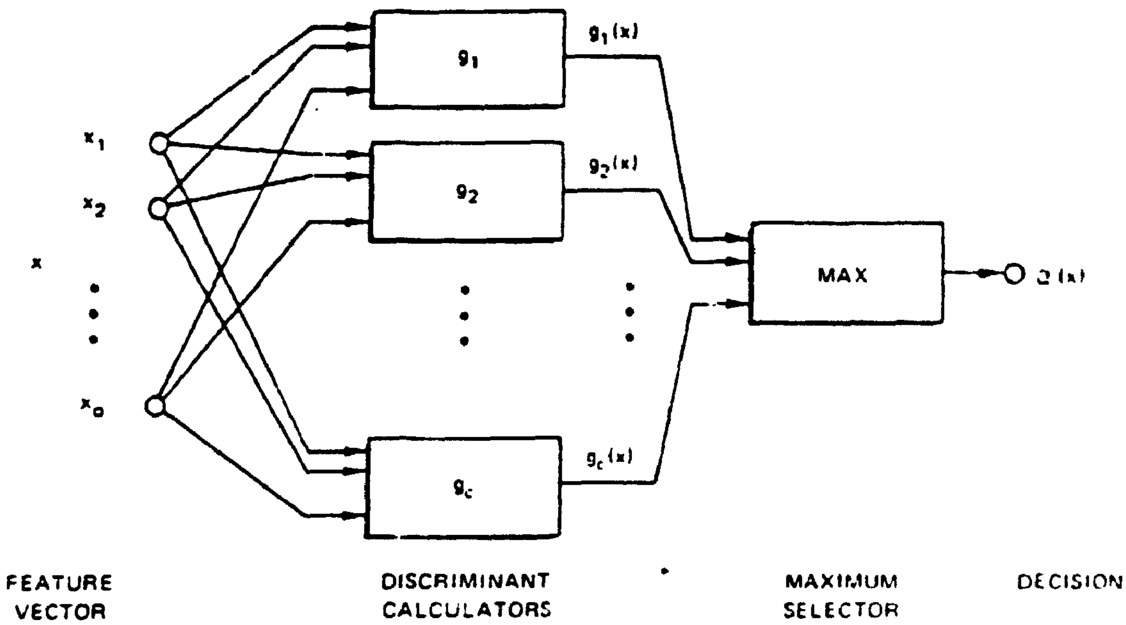


Figure 1.2: Diagram of pattern identification

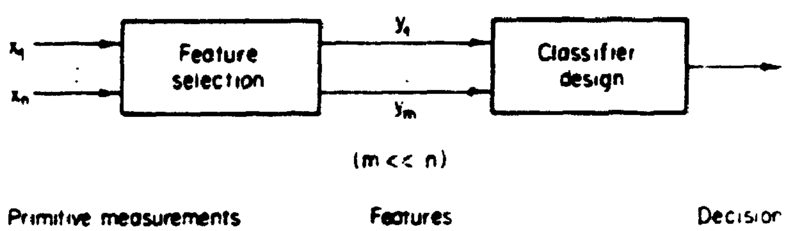


Figure 1.3: Diagram of feature selection

Another problem can occur when one or more variables do not make a significant contribution to the total variation because of the existence of other useful variables. This problem might be solved by employing the so-called *stepwise variable selection process* [Jennrich, 1977].

Therefore, point (a) above can be divided further into

(a.1) the determination of random variables in the measurement vector ;

(a.2) the feature selection process, including the variable selection process (see Figure 1.3).

One of the objectives of this research is to develop a generalized method of pattern recognition that is flexible and precise and makes optimal use of measurement data. We use cryptanalysis as a test bed for this research.

1.3 A Simple Example

Consider a sample of one hundred texts of similar type or writing style, for example articles in *Computing Canada*.

For each text we calculate the relative frequency of characters [A,B,C,D,....,X,Y,Z, and @ (space, blank)], and then take the average of the relative frequencies. Let say that the result is as follows :

Char	Rel.Freq	Char	Rel.Freq	Char	Rel.Freq
A	0.071292	J	0.001170	S	0.059985
B	0.011646	K	0.005394	T	0.074318
C	0.034395	L	0.033322	U	0.023310
D	0.031729	M	0.026321	V	0.009304
E	0.096478	N	0.063462	W	0.012535
F	0.017467	O	0.064976	X	0.002715
G	0.016217	P	0.023460	Y	0.013247
H	0.028782	Q	0.001143	Z	0.000875
I	0.062482	R	0.054609	@	0.159366

Table 1.1: Sample average of relative frequencies

This information can, then, be used to identify every character of a text from the same type, which have been encrypted, say, using the monoalphabetic substitution method. We can compare the relative frequency of every encrypted character with the relative frequency in the above table and make the decision accordingly, often involving our subjective judgment whether the translation makes sense or not. Why is subjective judgement required ? Because some characters have almost the same relative frequency.

Actually, we have just discussed a very simple discriminant analysis to identify the encrypted characters by using the table of relative frequency that we have made in advance. And this simple methodology is still used in cryptanalysis.

1.3.1 The components involved in the discriminant analysis

a. **FEATURE VECTOR** : in the previous example, it consists of a single variable (relative frequency).

b. **DISCRIMINANT CALCULATOR** : in the example given, it is the table of relative frequency of each character.

In the real situation of discriminant analysis, this calculator is built by using a sample of objects (in this case the text known previously). Every character will have its own function (usually one or more coefficients to be applied to the values of one or more variables, respectively). The results from this calculator are the discriminant scores of the characters (one for each character).

The next section will give an example of discriminant analysis based on a sample of two hundred texts using only one variable relative frequency of every character in the text. About fifty percent of the sample will be used as a training sample to build the discrim inant calculator, and another fifty percent will be used as a validating sample to test the performance of the calculator.

c. **MAXIMUM SELECTOR** : an algorithm to determine the minimum or the maximum value of the discriminant scores given by the discriminant calculator.

d. **DECISION**. The character that has the extreme value usually is chosen to identify an unknown or encrypted character.

1.3.2 Example of discriminant calculator using only one variable (relative frequency of character)

The functions in Table 1.2 are the results of the linear discriminant analysis using a training sample of 2674 measurement vectors, which consists of one variable [RELFREQ]. The explanation of this analysis is given in Chapter 3. The coefficients and the constants can be used

Table 1.2: Coefficients of linear discrimination for character groups by using one variable

Group	Coefficients of RELFREQ	CONSTANT
A	2451.34766	-88.34501
B	427.62061	-7.06080
C	1247.35107	-25.56044
D	1107.85669	-20.95539
E	3437.08569	-170.82770
F	642.20062	-9.92955
G	527.38367	-8.08854
H	1000.28503	-17.81850
I	2162.29590	-69.45689
J	37.09273	-6.77037
K	187.06538	-5.72156
L	1156.78723	-22.48681
M	931.46771	-16.01187
N	2227.84033	-73.54531
O	2281.71167	-76.98656
P	835.21796	-13.70173
Q	47.97626	-6.80693
R	1933.44788	-56.22342
S	2099.72778	-65.69437
T	2612.28491	-99.92624
U	807.90601	-13.06809
V	322.35156	-6.15932
W	427.76773	-6.98903
X	109.97643	-6.08146
Y	442.89038	-7.12157
Z	23.69461	-7.04929
@	5575.79346	-445.24588

to classify a measurement vector which consists of only one value of variable RELFREQ into one of the character groups. The value of RELFREQ in the measurement vector is multiplied by the coefficient and is then added to the constant of every group. In this way, we will obtain twenty seven scores, one for each character type. The maximum of these scores indicates that the measurement vector belongs to the character type that has the maximum score.

The application of these coefficients and constants to the measurement vectors in the training and validating samples produces the classification matrices given in Tables 1.3 and 1.4.

We can observe from the classification matrices that there are a lot of misclassifications among characters. For examples, some of C's are misclassified as D's, and vice versa ; some C's and D's are misclassified as L's and M's, and vice versa. Only character @ (blank/space) is correctly classified one hundred percent. The total correctness of all characters is only 26.81 percent in the training sample and 27.51 percent in the validating sample.

To understand this situation, we can check the sample distribution of every character. In Figure 1.4, we can see that the distribution of characters C, D, L, and M are overlapping to each other. Character @ is exclusively separated from other characters. In other words, the variable Relative Frequency is not good enough to discriminate between characters.

Table 1.3: Classification matrix for the training sample using only one variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.63	31	0	0	0	2	0	0	0	5	0	0	0	0	5
B	0.00	0	0	0	0	0	5	48	0	0	0	0	0	0	0
C	50.50	0	0	51	7	0	0	0	9	0	0	0	17	5	0
D	20.21	0	0	24	19	0	0	0	15	0	0	0	22	8	0
E	92.66	0	0	0	0	101	0	0	0	0	0	0	0	0	0
F	34.48	0	0	0	0	0	30	20	1	0	0	0	0	3	0
G	49.48	0	0	0	0	0	23	48	1	0	0	0	0	1	0
H	19.39	0	0	20	13	0	2	1	19	0	0	0	11	14	0
I	11.36	8	0	0	0	0	0	0	0	10	0	0	0	0	9
J	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
K	13.19	0	0	0	0	0	0	1	0	0	0	12	0	0	0
L	24.24	0	0	33	19	0	0	0	14	0	0	0	24	2	0
M	12.77	0	0	11	6	0	7	1	17	0	0	0	5	12	0
N	11.43	19	0	0	0	0	0	0	0	8	0	0	0	0	12
O	19.05	26	0	1	0	0	0	0	0	9	0	0	0	0	10
P	5.88	0	0	3	5	0	5	5	13	0	0	0	3	22	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
R	62.63	0	0	5	0	0	0	0	0	9	0	0	0	0	7
S	11.88	12	0	4	0	0	0	0	0	12	0	0	0	0	8
T	52.75	22	0	0	0	9	0	0	0	3	0	0	0	0	2
U	30.63	0	0	1	3	0	13	6	15	0	0	0	1	24	0
V	30.77	0	0	0	0	0	2	19	0	0	0	1	0	0	0
W	4.30	0	0	0	0	0	6	40	0	0	0	0	0	1	0
X	0.00	0	0	0	0	0	0	1	0	0	0	63	0	0	0
Y	28.71	0	0	0	0	0	12	49	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	101	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	26.81	118	0	153	72	112	105	239	104	56	0	369	83	92	53

Table 1.3 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	31.63	19	0	0	1	5	30	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	2	8	2	0	31	0	0	96
C	50.50	0	2	0	9	0	0	1	0	0	0	0	0	0	101
D	20.21	0	1	0	2	0	0	3	0	0	0	0	0	0	94
E	92.66	0	0	0	0	0	8	0	0	0	0	0	0	0	109
F	34.48	0	1	0	0	0	0	31	0	0	0	1	0	0	87
G	49.48	0	0	0	0	0	0	10	2	0	0	12	0	0	97
H	19.39	0	3	0	0	0	0	14	0	1	0	0	0	0	98
I	11.36	16	0	0	20	17	8	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	4	0	0	0	0	0	100
K	13.19	0	0	0	0	0	0	1	69	3	0	5	0	0	91
L	24.24	0	1	0	4	0	0	2	0	0	0	0	0	0	99
M	12.77	0	11	0	2	0	0	22	0	0	0	0	0	0	94
N	11.43	19	0	0	19	17	11	0	0	0	0	0	0	0	105
O	19.05	20	0	0	12	11	16	0	0	0	0	0	0	0	105
P	5.88	0	6	0	0	0	0	40	0	0	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	6	0	0	0	0	0	102
R	62.63	3	0	0	62	12	1	0	0	0	0	0	0	0	99
S	11.88	11	0	0	34	12	8	0	0	0	0	0	0	0	101
T	52.75	4	0	0	2	1	48	0	0	0	0	0	0	0	91
U	30.63	0	14	0	0	0	0	34	0	0	0	0	0	0	111
V	30.77	0	0	0	0	0	0	0	32	14	0	36	0	0	104
W	4.30	0	0	0	0	0	0	3	7	4	0	32	0	0	93
X	0.00	0	0	0	0	0	0	0	40	1	0	0	0	0	105
Y	28.71	0	0	0	0	0	0	2	3	6	0	29	0	0	101
Z	0.00	0	0	0	0	0	0	0	1	0	0	0	0	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
Total	26.81	92	39	0	167	75	130	165	172	31	0	146	0	101	2674

Table 1.4: Classification matrix for the validating sample using only one variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.37	32	0	0	0	1	0	0	0	6	0	0	0	0	5
B	0.00	0	0	0	0	0	6	40	0	0	0	0	0	1	0
C	47.47	0	0	47	11	0	0	0	10	0	0	0	13	9	0
D	13.21	0	0	37	14	0	0	0	20	0	0	0	20	9	0
E	93.41	0	0	0	0	85	0	0	0	0	0	0	0	0	0
F	33.63	0	0	0	2	0	38	30	1	0	0	0	0	6	0
G	47.57	0	0	0	0	0	26	49	0	0	0	0	0	1	0
H	12.75	0	0	18	15	0	2	0	13	0	0	0	20	15	0
I	12.50	15	0	0	0	0	0	0	0	14	0	0	0	0	8
J	0.00	0	0	0	0	0	0	0	0	0	0	98	0	0	0
K	22.94	0	0	0	0	0	0	1	0	0	0	25	0	0	0
L	17.82	0	0	38	17	0	0	0	14	0	0	0	18	7	0
M	20.75	0	0	5	13	0	5	1	19	0	0	0	9	22	0
N	8.42	13	0	0	0	1	0	0	0	15	0	0	0	0	8
O	23.16	20	0	0	0	0	0	0	0	12	0	0	0	0	15
P	9.18	0	0	3	5	0	11	5	8	0	0	0	1	18	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	90	0	0	0
R	51.49	1	0	4	0	0	0	0	0	8	0	0	0	0	8
S	14.14	12	0	0	0	0	0	0	0	13	0	0	0	0	8
T	68.81	17	0	0	0	6	0	0	0	1	0	0	0	0	1
U	34.83	0	0	0	2	0	7	5	17	0	0	0	4	18	0
V	47.92	0	0	0	0	0	0	7	0	0	0	1	0	0	0
W	10.28	0	0	0	0	0	14	45	0	0	0	0	0	0	0
X	0.00	0	0	0	0	0	0	0	0	0	0	65	0	0	0
Y	26.26	0	0	0	0	0	9	53	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	97	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	27.51	110	0	152	79	93	118	236	102	69	0	376	85	106	53

Table 1.4 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	31.37	14	0	0	0	2	42	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	4	8	4	0	41	0	0	104
C	47.47	0	1	0	5	0	0	3	0	0	0	0	0	0	99
D	13.21	0	0	0	1	0	0	5	0	0	0	0	0	0	106
E	93.41	0	0	0	0	0	6	0	0	0	0	0	0	0	91
F	33.63	0	0	0	0	0	0	33	0	0	0	3	0	0	113
G	47.57	0	1	0	0	0	0	14	0	0	0	12	0	0	103
H	12.75	0	2	0	1	0	0	16	0	0	0	0	0	0	102
I	12.50	18	0	0	33	15	9	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	2	0	0	0	0	0	100
K	22.94	0	0	0	0	0	0	0	67	10	0	6	0	0	109
L	17.82	0	1	0	4	0	0	2	0	0	0	0	0	0	101
M	20.75	0	11	0	0	0	0	21	0	0	0	0	0	0	106
N	8.42	7	0	0	28	14	9	0	0	0	0	0	0	0	95
O	23.16	22	0	0	6	10	10	0	0	0	0	0	0	0	95
P	9.18	0	9	0	0	0	0	38	0	0	0	0	0	0	98
Q	0.00	0	0	0	0	0	0	0	8	0	0	0	0	0	98
R	51.49	8	0	0	52	20	0	0	0	0	0	0	0	0	101
S	14.14	24	0	0	26	14	2	0	0	0	0	0	0	0	99
T	68.81	7	0	0	1	1	75	0	0	0	0	0	0	0	109
U	34.83	0	5	0	0	0	0	31	0	0	0	0	0	0	89
V	47.92	0	0	0	0	0	0	0	46	10	0	32	0	0	96
W	10.28	0	0	0	0	0	0	2	7	11	0	28	0	0	107
X	0.00	0	0	0	0	0	0	0	29	0	0	1	0	0	95
Y	26.26	0	0	0	0	0	0	3	5	3	0	26	0	0	99
Z	0.00	0	0	0	0	0	0	0	0	1	0	0	0	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
Total	27.51	100	30	0	157	76	153	172	172	39	0	149	0	99	2726

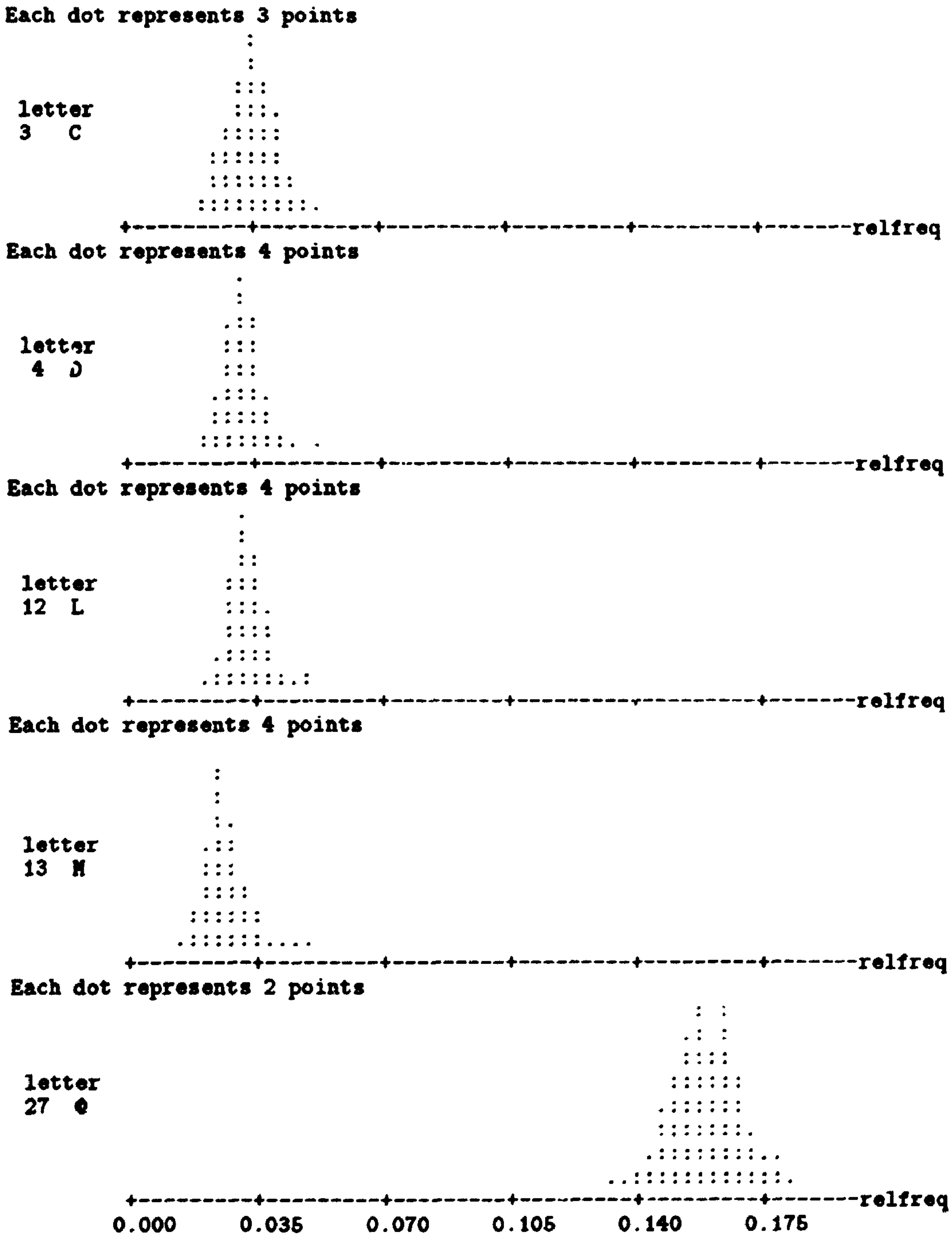


Figure 1.4: Sample distribution of relative frequency for some characters

1.4 Identification of Letters in English Texts

1.4.1 Monoalphabetic substitution problem

Figure 1.5 shows a sample of *source text* in English which consists of twenty seven groups of capital letters A, B, C,$\cdots$ and Blank [Konheim, 1981]. Special characters - such as period, comma, and the like - and numbers have been erased from the text. If we substitute the letters in the source text using the rules in Table 1.5 then we will obtain the *crypto text*

A --> Blank	J --> R	S --> I
B --> Z	K --> Q	T --> H
C --> Y	L --> P	U --> G
D --> X	M --> O	V --> F
E --> W	N --> N	W --> E
F --> V	O --> M	X --> D
G --> U	P --> L	Y --> C
H --> T	Q --> K	Z --> B
I --> S	R --> J	Blank --> A

Table 1.5: One of the rules of monoalphabetic substitution

shown in Figure 1.6. Information for each group of letters in the texts can be seen in Table 1.6.

The main problem in *monoalphabetic substitution* is to identify letters in the crypto text which substitute for letters in the source text in such a way that one can regain the meaningful message of the source text.

One of the methods that can be used is the *exhaustive method*, in which one examines all of the possibilities of substitution. But, one should be aware that the number of possibilities is in the order of $27!$ ($\approx 1.089 \times 10^{28}$).

Another method is by taking a sample of source text and calculating the sample average of a variable representing the relative frequency of every group of letters. Then, one can calculate the relative frequencies of every group of letters in the crypto text and compare them

---

MANY ORGANIZATIONS RELY ON COMPUTERS AND DATA COMMUNICATIONS
 TO KEEP THEIR OPERATIONS RUNNING SMOOTHLY BY MAKING INFORMA
TION MORE ACCESSIBLE TO MORE PEOPLE WITHIN THE ORGANIZATION
BUT AS IT BECOMES MORE ACCESSIBLE INFORMATION REQUIRES MORE
PROTECTION THAN YOU MAY NOW HAVE YOU CAN NOW PROTECT INFORMA
TION STORED ON YOUR PREMISES BY PHYSICAL MEASURES THAT LIMIT
 ACCESS TO AUTHORIZED PEOPLE SIMILARLY IBM HARDWARE AND SOFT
WARE PRODUCTS HAVE FEATURES THAT CAN BE USED TO IDENTIFY AND
 CHECK THE AUTHORIZATION OF PEOPLE TRYING TO GAIN ACCESS TO
A SYSTEM AND ITS INFORMATION NOW THERE IS A WAY TO PROTECT I
NFORMATION EVEN FURTHER THE IBM CRYPTOGRAPHIC SUBSYSTEM CAN
EXTEND DATA CONTROL AND PROTECTION TO THE DATA COMMUNICATION
S TERMINALS AND LINKS THAT SPEED INFORMATION FROM ONE LOCATI
ON TO ANOTHER IT USES A SOPHISTICATED ALGORITHM A STRICT SET
 OF RULES TO ENCRYPT OR SCRAMBLE DATA BEFORE IT IS STORED OR
 TRANSMITTED TO ANOTHER LOCATION AND DECRYPT IT WHEN NEEDED
FOR PROCESSING IT EMPLOYS ENCRYPTION TECHNIQUES THAT CAN RED
UCE INFORMATION EXPOSURES WITHIN YOUR COMMUNICATIONS NETWORK
 AS WELL AS PROVIDE A SYSTEM BASE FOR THE DEVELOPMENT OF ENC
RYPTION PROGRAMS THE IBM CRYPTOGRAPHIC SUBSYSTEM IS A VERSAT
ILE TOOL FOR CONTROLLING AND PROTECTING INFORMATION THROUGH
ENCRYPTION BY A COMBINATION OF PROGRAMMING AND SNA TERMINAL
HARDWARE FEATURES IT CAN ENCRYPT AND DECRYPT INFORMATION AUT
OMATICALLY AND WITHOUT INTERVENTION BY THE TERMINAL USER OR
APPLICATION USING AN ALGORITHM AND A KEY WHICH INDIVIDUALIZE
S THE ALGORITHM THE SUBSYSTEM ENCRYPTS APPLICATION INFORMATI
ON BEFORE IT IS SENT FROM A TERMINAL OR COMPUTER LOCATION AN
D ENTERS YOUR DATA COMMUNICATIONS NETWORK AT THE RECEIVING T
ERMINAL OR COMPUTER LOCATION THE SAME KEY IS USED TO DECRYPT
 THE INFORMATION AFTER IT LEAVES THE NETWORK IN ADDITION TO
THE ALGORITHM THE IBM SUBSYSTEM PROVIDES KEY GENERATION KEY
MANAGEMENT VERIFICATION AND OPERATIONAL FEATURES THAT ENHANC
E THE BASIC CRYPTOGRAPHIC SECURITY OF THE SUBSYSTEM

Figure 1.5: A sample of source text

```
---------------------------------------------------------------

O NCAMJU NSB HSMNIAJWPCAMNAYMOLGHWJIA NXAX H AYMOOGNSY HSMNI
AHMAQWWLAHTWSJAMLWJ HSMNIAJGNNSNUAIOMMHTPCAZCAO QSNUASNVMJO
HSMNAOMJWA YYWIISZPWAHMAOMJWALWMLPWAESHTSNAHTWAMJU NSB HSMNA
ZGHA IASHAZWYMOWIAOMJWA YYWIISZPWASNVMJO HSMNAJWKGSJWIAOMJWA
LJMHWYHSMNAHT NACMGAO CANMEAT FWACMGAY NANMEALJMHWYHASNVMJO
HSMNAIHMJWXAMNACMGJALJWOSIWIAZCALTCISY PAOW IGJWIAHT HAPSOSH
A YYWIIAHMA GHTMJSBWXALWMLPWAISOSP JPCASZOAT JXE JWA NXAIMVH
E JWALJMXGYHIAT FWAVW HGJWIAHT HAY NAZWAGIWXAHMASXWNHSVCA NX
AYTWYQAHTWA GHTMJSB HSMNAMVALWMLPWAHJCSNUAHMAU SNA YYWIIAHMA
 AICIHWOA NXASHIASNVMJO HSMNANMEAHTWJWASIA AE CAHMALJMHWYHAS
NVMJO HSMNAWFWNAVGJHTWJAHTWASZOAYJCLHMUJ LTSYAIGZICIHWOAY NA
WDHWNXAX H AYMNHJMPA NXALJMHWYHSMNAHMAHTWAX H AYMOOGNSY HSMN
IAHWJOSN PIA NXAPSNQIAHT HAILWWXASNVMJO HSMNAVJMOAMNWAPMY HS
MNAHMA NMHTWJASHAGIWIA AIMLTSIHSY HWXA PUMJSHTOA AIHJSYHAIWH
AMVAJGPWIAHMAWNYJCLHAMJAIYJ OZPWAX H AZWVMJWASHASIAIHMJWXAMJ
AHJ NIOSHHWXAHMA NMHTWJAPMY HSMNA NXAXWYJCLHASHAETWNANWWXWXA
VMJALJMYWIISNUASHAWOLPMCIAWNYJCLHSMNAHWYTNSKGWIAHT HAY NAJWX
GYWASNVMJO HSMNAWDLMIGJWIAESHTSNACMGJAYMOOGNSY HSMNIANWHEMJQ
A IAEWPPA IALJMFSXWA AICIHWOAZ IWAVMJAHTWAXWFWPMLOWNHAMVAWNY
JCLHSMNALJMUJ OIAHTWASZOAYJCLHMUJ LTSYAIGZICIHWOASIA AFWJI H
SPWAHMMPAVMJAYMNHJMPPSNUA NXALJMHWYHSNUASNVMJO HSMNAHTJMGUTA
WNYJCLHSMNAZCA AYMOZSN HSMNAMVALJMUJ OOSNUA NXAIN AHWJOSN PA
T JXE JWAVW HGJWIASHAY NAWNYJCLHA NXAXWYJCLHASNVMJO HSMNA GH
MO HSY PPCA NXAESHTMGHASNHWJFWNHSMNAZCAHTWAHWJOSN PAGIWJAMJA
 LLPSY HSMNAGISNUA NA PUMJSHTOA NXA AQWCAETSYTASNXSFSXG PSBW
IAHTWA PUMJSHTOAHTWAIGZICIHWOAWNYJCLHIA LLPSY HSMNASNVMJO HS
MNAZWVMJWASHASIAIWNHAVJMOA AHWJOSN PAMJAYMOLGHWJAPMY HSMNA N
XAWNHWJIACMGJAX H AYMOOGNSY HSMNIANWHEMJQA HAHTWAJWYWSFSNUAH
WJOSN PAMJAYMOLGHWJAPMY HSMNAHTWAI OWAQWCASIAGIWXAHMAXWYJCLH
AHTWASNVMJO HSMNA VHWJASHAPW FWIAHTWANWHEMJQASNA XXSHSMNAHMA
HTWA PUMJSHTOAHTWASZOAIGZICIHWOALJMFSXWIAQWCAUWNWJ HSMNAQWCA
O N UWOWNHAFWJSVSY HSMNA NXAMLWJ HSMN PAVW HGJWIAHT HAWNT NY
WAHTWAZ ISYAYJCLHMUJ LTSYAIWYGJSHCAMVAHTWAIGZICIHWO

---------------------------------------------------------------
```

Figure 1.6: Crypto text from the sample text after monoalphabetic substitution

Source text letters	Frequency	Relative Frequency	Crypto text letters
A	141	0.071537	Blank
B	24	0.012177	Z
C	71	0.036022	Y
D	45	0.022831	X
E	166	0.084221	W
F	31	0.015728	V
G	24	0.012177	U
H	55	0.027905	T
I	138	0.070015	S
J	0	0.000000	R
K	11	0.005581	Q
L	43	0.021816	P
M	69	0.035008	O
N	135	0.068493	N
O	155	0.078640	M
P	49	0.024860	L
Q	2	0.001015	K
R	121	0.061390	J
S	94	0.047692	I
T	178	0.090309	H
U	42	0.021309	G
V	12	0.006088	F
W	16	0.008118	E
X	2	0.001015	D
Y	44	0.022324	C
Z	5	0.002537	B
Blank	298	0.151192	A
Total	1971	1.000000	

Table 1.6: Information for each group of letters

with the averages obtained from the sample of source text. This is possible since the variable relative frequency has an important feature called *invariance*, so that one can use this feature to compare both texts. This *univariate method* – which uses only one variable, i.e. relative frequency – has been used for a long time as the basis for cryptanalysis (the study of breaking the secret of crypto texts) [Bosworth,1963][Yardley,1981][Robbins,1988][Gaines,1956] [Smith,1955][Tannenbaum,1981][Foster,1990].

Problem 1 : By using only the sample averages of relative frequencies, some information about the distribution of sample data becomes unknown. In general, one cannot use the variation contained in this variable, namely between-groups and within-group variations. They are very useful to build a method of discriminating and classifying groups of objects as mentioned in the previous section.

Problem 2 : If the single variable representing relative frequency does not provide enough variation to uniquely identify all of the letters, one will need to generate other invariant variables that can contribute to the total variation. The question is : how do we generate these variables from the sample of source texts ?

Problem 3 : How much is the contribution of each of the variables to the total variation; and to the discriminating power of some method of identification.

1.4.2 Polyalphabetic substitution problem

Solving polyalphabetic substitution ciphers is a more difficult problem, which we shall investigate to challenge the power of our pattern classification technique.

Let consider the sample source text in Figure 1.5 as a sequence of strings of three characters. If we substitute each character in every position of the strings with the rules in Table 1.7 then we will get the crypto text given in Figure 1.7. Information for each group of letters in the texts can be seen in Table 1.8.

Problem 4 : The tables show that the variable, relative frequency, is no longer invariant since there is no one-to-one mapping between the letters in source text and crypto text. This is the strength of polyalphabetic substitution compared to monoalphabetic

```
------------------------------------------------------------
123123123123123123123123123123123123123123123123123123123123
------------------------------------------------------------
PULATMU ZQBXDMGRGQCKCORYRGYFHKSNRHKQCULGTBDMZCWMPFSQBADMGRGQ
CMMCDCHIYWACLKYRICUURLHLVTPXGLLGECLKRHRKEWCVWCFZNBLJTGQZMUFZ
WBMQTKRKCCUAFYQVB OYYWHYPHPHTNHHNOYYZBRKBLCMFHTMU ZQBXDMGRGY
ENRCUQCBRCVCFHKHLYPHPHTZFWCVLGEECCBLIHPPURLHLCKCTNGUYQCFMUYY
SKMWYAWBMQTRKULCRMXTKDRYQHUCAZYYYAHSCWZQTLRPYSKMWYAWTGQZMUFZ
WBMQTQWHPHXYRGYAHSUTNUYKLLCVT ATNKRQLWZOTKHUQXKCVTRKURCEGPBR
CUAFYQVTRRTZXMFRKGBYBCICRIJHTQLFGOUPORYLVKCAZUXUDKCCULGTQRZR
ZUPHTNUHBXWRVTFDOCCZCDMSUYQCMFDMYFULCVCCNQHXYWHYLXCQMGIRYDGB
CWFHWICMFHTZXMFRKGBURLHLCHDCICRIJHTRURGQ YWHYJUGQTZFWCVLYWHY
DTQALRHFYDGBCBRVTGQZMUFZWBMQTLRPYWACUYYLLYDTUDRYWHYSKMWYAWTG
QZMUFZWBMQTCYYLCZSUMFHKYWACCB PTAURNWHEUUNKBACLSELWVMCPTADGY
HQRHGBCXZWUYFHLWKMOTZQXYSKMWYAWBMQTRRTRKYYGURDTARFKXGGFURLHL
VTRHKKLGZOLYDGBCEGQDQCMFDMYVICHXYLGDRKKDMGRGYIKMPTMQYYOHADMG
RGYWHYDGMWACUTGWTSVYQCUYVHNKBQWBADMCGTZO MUBRKFYDTQWKGFMYVYR
CHDCKSOYQCMMCYLFKWSMYRKYVWPDF OYYGURDT HZMUYYLMYLLYVMMUYBCHP
CMPDGQPBRWYBCMMCULRMFHKYOHADMGRGYDGBCXCFKWSMYLMYZACQTLHYBHXY
IHPCIPRWCVLGQ YLMYHFNOHWVTCQWPAIRLHLCMCFALLJSHLYWAZWTADGYUYB
XWCCBLIHPPURLHLCYVSHQXKCVTULMFLGYAHSUTARFKXGGFURLHLVTLHMURKI
CUQCPCOEYDLYSKMYBBHTZCLWVMCPT DLCCZMUTRKYYGYTHEMSFCQMYRZYHGA
URNWBMQTNUHEUUKVTRKYYLVKCWPAIRR PDIFLWYVN VRQWYKCBQCUYYYPVUR
LECCMMREYIHPCWMQMPREJLGECULGTNUHRHWRLGECBLIHPPURLHLCMFUHSJAY
HGAURNWBMQT ATZCWMPVGQURLHLCHDCIPR PDFKLGECULGTQQUYWYPPBLDEY
KUPGPZUYYIYZWNPHLYLMYFULCYLFKWSMYDGBCXCFKWSMYLGDRKKDMGRGYDNR
RFZWBADEJATZQXYZBRKHSWTGQMCUOCQMGRGYERYWACCMCUFGQUJCNQHKYRKY
DINOBADMGRGYXLGQ YDGYDEERKGWAKCULGTZCDCATUKBAKTGQXGYBBXUJLSC
VTRKYYDEERKGWAKCMFHTQXVQALRHFYHGAURNWLYDINOBADMGRGYLGDRKKDMG
RGYEYDRKCCBRCBQCLCQMYIKMPTZCMCUFGQUJCHPCWMPISWYPCEMFURLHLCUL
GTCQMCULYAHSUTBDMZCWMPFSQBADMGRGQCGCWPMUDYDMYWACCKCFYGYBLJTR
HKKLGZOTMUTARFNXMCUTJRWZWBMQTRKYYVUKHTIHRYLLYXLCGTRRTBHWPAIR
CMFHTGQZMUFZWBMQTZIMCUTGWTJHUTHLYWACCGCWPMUDYLGYDXBLMGRGYWHY
WACCUJJHPLMFPTRKYYLVKCLSELWVMCPTNUHTLXCVTIHRYJYLHKZWBMQTIHRY
PULD CPYLWTTHKGIBADMGRGYDGBCHNHKZWBMQUJCZCDMSUYQCMFDMYHGFDGA
HTRKYYEUQLWYFKWSMMJKZSAGFTQHWSUBRATMITRKYYVN VRQWYK
------------------------------------------------------------
```

Figure 1.7: Crypto text from the sample text after polyalphabetic substitution

Position 1		Position 2		Position 3	
A -> D	O -> R	A -> U	O -> H	A -> Z	O -> M
B -> E	P -> S	B -> V	P -> I	B -> @	P -> N
C -> F	Q -> T	C -> W	Q -> J	C -> A	Q -> O
D -> G	R -> U	D -> X	R -> K	D -> B	R -> P
E -> H	S -> V	E -> Y	S -> L	E -> C	S -> Q
F -> I	T -> W	F -> Z	T -> M	F -> D	T -> R
G -> J	U -> X	G -> @	U -> N	G -> E	U -> S
H -> K	V -> Y	H -> A	V -> O	H -> F	V -> T
I -> L	W -> Z	I -> B	W -> P	I -> G	W -> U
J -> M	X -> @	J -> C	X -> Q	J -> H	X -> V
K -> N	Y -> A	K -> D	Y -> R	K -> I	Y -> W
L -> O	Z -> B	L -> E	Z -> S	L -> J	Z -> X
M -> P	@ -> C	M -> F	@ -> T	M -> K	@ -> Y
N -> Q		N -> G		N -> L	

Note : @ represents blank or space

Table 1.7: One of the rules of polyalphabetic substitution with key length = 3

substitution. As in monoalphabetic substitution, one can solve the problem in polyalphabetic substitution by using *exhaustive method*. But, once again, the number of possibilities of substituting characters is very large number, i.e., $27! \times 27! \times 27!$ when key length = 3.

1.5 Research objectives

My thesis is that the statistical analysis of variation can be applied to the general problem of pattern recognition to resolve ambiguities that arise when parts of the pattern are obscured or aliased.

A worst-case situation arises in the automated cryptanalysis of polyalphabetic substitution ciphers. Here the groups of objects are the symbols of the plain-text alphabet. In encrypted text, the symbols can be called by several different names depending upon the key to the cipher.

Traditionally, the solution has depended upon properly classifying the symbols found in the encrypted text into their corresponding plain-text groups according to their mean

Source letters	Freq.	Relative Frequency	Crypto letters	Freq.	Relative Frequency
A	141	0.071537	A	59	0.029934
B	24	0.012177	B	67	0.033993
C	71	0.036022	C	157	0.079655
D	45	0.022831	D	68	0.034500
E	166	0.084221	E	29	0.014713
F	31	0.015728	F	62	0.031456
G	24	0.012177	G	108	0.054795
H	55	0.027905	H	107	0.054287
I	138	0.070015	I	33	0.016743
J	0	0.000000	J	19	0.009640
K	11	0.005581	K	88	0.044647
L	43	0.021816	L	113	0.057331
M	69	0.035008	M	119	0.060375
N	135	0.068493	N	29	0.014713
O	155	0.078640	O	20	0.010147
P	49	0.024860	P	57	0.028919
Q	2	0.001015	Q	81	0.041096
R	121	0.061390	R	121	0.061390
S	94	0.047692	S	33	0.016743
T	178	0.090309	T	97	0.049214
U	42	0.021309	U	102	0.051750
V	12	0.006088	V	41	0.020802
W	16	0.008118	W	91	0.046169
X	2	0.001015	X	31	0.015728
Y	44	0.022324	Y	168	0.085236
Z	5	0.002537	Z	53	0.026890
@	298	0.151192	@	18	0.009132
Total	1971	1.000000	Total	1971	1.000000

Table 1.8: Information for each letter type

occurence or co-occurence frequencies. This can be regarded as using first moment information. The analysis has been rule-based and has relied upon expected occurence frequencies of digraphs in normal text [Carroll *et al*, 1986].

Statistical variation analysis will permit a methodological approach to classification and the use of exhausting statistical information about digraphs, tri-graphs and n-graphs. It will be shown that this will make a quantitative improvement in the time, precision and reliability of automated cryptanalysis of polyalphabetic substitution ciphers in exchange for greater computational effort and memory thus pointing the way to resolving uncertainty in the generalized pattern-recognition problem.

Following the model given in Figure 1.2 and by taking into account the problems mentioned in this chapter, the objectives of this research are :

(1) to utilize variation as the basis for identifying groups of letters in crypto texts through discrimination and classification processes using a sample of source texts ;

(2) to increase the amount of information and variation by generating more variables - which have the *invariance* between source text and crypto text – from the sample of source texts. In univariate methods, the variable *relative frequency* is usually obtained by utilizing 1-graph structure of source texts. For multivariate methods, more variables can be generated by using digraph structure or higher structure of source texts. This research will use the digraph structure to generate more variables. Chapter 2 discusses this matter in detail ;

(3) to use a method in Multivariate Statistical Analysis called Multivariate Discriminant Analysis to discriminate and to classify groups of letters in texts. This method will be discussed in Chapter 3 ;

(4) to compare the results of Discriminant Analysis when applied to both univariate and multivariate situations : how much improvement is there in identifying groups of letter by using the multivariate method compared to univariate method ?

CHAPTER 2

STRUCTURES OF STRINGS OF CHARACTERS

2.1 Introduction

Consider two different strings of twenty characters from alphabet {A, B, C, D} as follows :

1. DCDDDACBDCCDDACCBBBD ;

2. ACDADBCCDDCDCDDDBCBB .

Information about every character, based on 1-graph structure, can be seen in Table 2.1

Table 2.1: Frequency and Relative Frequency of Characters from Different Strings Based on 1-graph Structure

1-graph	String-1		String-2	
	Frequency	Relative Frequency	Frequency	Relative Frequency
A	2	0.10	2	0.10
B	4	0.20	4	0.20
C	6	0.30	6	0.30
D	8	0.40	8	0.40
Total	20	1.00	20	1.00

One can observe that the different strings have the same information based only on 1-graph structure.

The question is how to obtain different information from the different strings. To answer this problem we should try to use a different structure, which is called the 2-graph structure. We study all overlapping pairs, with the following rule be applied to the strings :

String-1

D C D D D A C B D C C D D A C C B B B D

String-2

A C D A D B C C D D C D C D D D B C B B

Information about every pair of characters, based on the 2-graph structure, can be seen in two different tables : Table 2.2 shows the information when a particular character is followed by any other character and Table 2.3 shows the information when a particular character follows any other character of the alphabet.

Furthermore, one can simplify the two Tables more concisely into two-dimensional arrays, for frequency (see Tables 2.4 and 2.5) and for relative frequency (see Tables 2.6 and 2.7).

Now, by observing and comparing the tables for the two strings of characters, one can see the different information between the two different strings. The marginal values, both in Column Total and in Row Total, are the same as the values in Table 2.1 , which are the values that come from 1-graph structure. In other words, a 2-graph structure presents more information than a 1-graph structure.

Table 2.2: Information available when a particular character is followed by any other character in 2-graph structure

Char 1	followed by	Char 2	2-graph	String-1 Freq.	String-1 Relative Freq.	String-2 Freq	String-2 Relative Freq.
A		A	AA	0	0.00	0	0.00
		B	AB	0	0.00	0	0.00
		C	AC	2	0.10	1	0.05
		D	AD	0	0.00	1	0.05
Subtotal for A			(A.)	2	0.10	2	0.10
B		A	BA	0	0.00	1	0.05
		B	BB	2	0.10	1	0.05
		C	BC	0	0.00	2	0.10
		D	BD	2	0.10	0	0.00
Subtotal for B			(B.)	4	0.20	4	0.20
C		A	CA	0	0.00	0	0.00
		B	CB	2	0.10	1	0.05
		C	CC	2	0.10	1	0.05
		D	CD	2	0.10	4	0.20
Subtotal for C			(C.)	6	0.30	6	0.30
D		A	DA	2	0.10	1	0.05
		B	DB	0	0.00	2	0.10
		C	DC	2	0.10	2	0.10
		D	DD	4	0.20	3	0.15
Subtotal for D			(D.)	8	0.40	8	0.40
TOTAL			(..)	20	1.00	20	1.00

The same situation can be seen from Tables 2.6 and 2.7, since relative frequency is just the ratio between the frequency and the total number of characters in the strings.

Higher structures, such as 3-graph and 4-graph, can be developed using the same procedure. 3-graph structure will result in 3-dimensional arrays with 4 x 4 x 4 cells, and 4 graph structures will produce 4-dimensional arrays with 4 x 4 x 4 x 4 cells.

In this research, the structure used is limited to 2-graph structures. We shall see the improvement of identifying characters using a 2-graph structure in comparison with the use of a 1-graph structure when the strings are transformed or encrypted.

Relative Frequency may be considered similar to the notion of probability, since they sum up to unity or 1.00 . Therefore, Tables 2.6 and 2.7 can be written or modified in terms

Table 2.3: Information available when a particular character follows any other character in 2-graph structure

Char 1 follows	Char 2	2-graph	String-1		String-2	
			Freq.	Relative Freq.	Freq	Relative Freq.
A	A	AA	0	0.00	0	0.00
	B	BA	0	0.00	1	0.05
	C	CA	0	0.00	0	0.00
	D	DA	2	0.10	1	0.05
Subtotal for A		(.A)	2	0.10	2	0.10
B	A	AB	0	0.00	0	0.00
	B	BB	2	0.10	1	0.05
	C	CB	2	0.10	1	0.05
	D	DB	0	0.10	2	0.10
Subtotal for B		(.B)	4	0.20	4	0.20
C	A	AC	2	0.10	1	0.05
	B	BC	0	0.00	2	0.10
	C	CC	2	0.10	1	0.05
	D	DC	2	0.10	2	0.10
Subtotal for C		(.C)	6	0.30	6	0.30
D	A	AD	0	0.00	1	0.05
	B	BD	2	0.10	0	0.00
	C	CD	2	0.10	4	0.20
	D	DD	4	0.20	3	0.15
Subtotal for D		(.D)	8	0.40	8	0.40
TOTAL		(..)	20	1.00	20	1.00

Table 2.4: Two-dimensional Array of Frequency for String-1

2-graph	A	B	C	D	Row Total
A	(AA) = 0	(AB) = 0	(AC) = 2	(AD) = 0	(A.) = 2
B	(BA) = 0	(BB) = 2	(BC) = 0	(BD) = 2	(B.) = 4
C	(CA) = 0	(CB) = 2	(CC) = 2	(CD) = 2	(C.) = 6
D	(DA) = 2	(DB) = 0	(DC) = 2	(DD) = 4	(D.) = 8
Column Total	(.A) = 2	(.B) = 4	(.C) = 6	(.D) = 8	Grand Total (..) = 20

Table 2.5: Two-dimensional Array of Frequency for String-2

2-graph	A	B	C	D	Row Total
A	(AA) = 0	(AB) = 0	(AC) = 1	(AD) = 1	(A.) = 2
B	(BA) = 1	(BB) = 1	(BC) = 2	(BD) = 0	(B.) = 4
C	(CA) = 0	(CB) = 1	(CC) = 1	(CD) = 4	(C.) = 6
D	(DA) = 1	(DB) = 2	(DC) = 2	(DD) = 3	(D.) = 8
Column Total	(.A) = 2	(.B) = 4	(.C) = 6	(.D) = 8	Grand Total (..) = 20

of 2-dimensional probability space. The first dimension of the probability space represents the occurence of a particular character in the first position of 2-graph structure; and the second dimension represents the occurence of a particular character in the second position. Table 2.8 illustrates this 2-dimensional probability space .

Table 2.6: Two-dimensional Array of Relative Frequency for String-1

2-graph	A	B	C	D	Row Total
A	(AA) = 0	(AB) = 0	(AC) = 0.10	(AD) = 0	(A.) = 0.10
B	(BA) = 0	(BB) = 0.10	(BC) = 0	(BD) = 0.10	(B.) = 0.20
C	(CA) = 0	(CB) = 0.10	(CC) = 0.10	(CD) = 0.10	(C.) = 0.30
D	(DA) = 0.10	(DB) = 0	(DC) = 0.10	(DD) = 0.20	(D.) = 0.40
Column Total	(.A) = 0.10	(.B) = 0.20	(.C) = 0.30	(.D) = 0.40	Grand Total (..) = 1.00

Table 2.7: Two-dimensional Array of Relative Frequency for String-2

2-graph	A	B	C	D	Row Total
A	(AA) = 0	(AB) = 0	(AC) = 0.05	(AD) = 0.05	(A.) = 0.10
B	(BA) = 0.05	(BB) = 0.05	(BC) = 0.10	(BD) = 0	(B.) = 0.20
C	(CA) = 0	(CB) = 0.05	(CC) = 0.05	(CD) = 0.20	(C.) = 0.30
D	(DA) = 0.05	(DB) = 0.10	(DC) = 0.10	(DD) = 0.15	(D.) = 0.40
Column Total	(.A) = 0.10	(.B) = 0.20	(.C) = 0.30	(.D) = 0.40	Grand Total (..) = 1.00

Table 2.8: Two-dimensional Probability Space from 2-graph Structure

	A	B	C	D	Row Total
A	p_{AA}	p_{AB}	p_{AC}	p_{AD}	$p_{A.}$
B	p_{BA}	p_{BB}	p_{BC}	p_{BD}	$p_{B.}$
C	p_{CA}	p_{CB}	p_{CC}	p_{CD}	$p_{C.}$
D	p_{DA}	p_{DB}	p_{CD}	p_{DD}	$p_{D.}$
Column Total	$p_{.A}$	$p_{.B}$	$p_{.C}$	$p_{.D}$	$p_{..} = 1.0$

If we denote A as 1, B as 2, C as 3, and D as 4 we will get the following table :

Table 2.9: Two-dimensional Probability Space from a 2-graph Structure (modified)

	1	2	3	4	Row Total
1	p_{11}	p_{12}	p_{13}	p_{14}	$p_{1.}$
2	p_{21}	p_{22}	p_{23}	p_{24}	$p_{2.}$
3	p_{31}	p_{32}	p_{33}	p_{34}	$p_{3.}$
4	p_{41}	p_{42}	p_{43}	p_{44}	$p_{4.}$
Column Total	$p_{.1}$	$p_{.2}$	$p_{.3}$	$p_{.4}$	$p_{.} = 1.0$

where

a. $p_{i.}$, $i = 1,2,3,4$ is the relative frequency/probability of the i-th character in the first position of 2-graph structure ;

b. $p_{.j}$, $j = 1,2,3,4$ is the relative frequency/probability of the j-th character in the second position of 2-graph structure ;

Notes : $p_{i.} = p_{i1} + p_{i2} + p_{i3} + p_{i4}$;
$p_{.j} = p_{1j} + p_{2j} + p_{3j} + p_{4j}$;

c. $p_{i.}$ = $p_{.i}, i = 1,2,3,4$, but it does not mean that $p_{ij} = p_{ji}$ for $j = 1,2,3,4$;
In 1-graph structure, $p_{i.}$ or $p_{.i}$ is the relative frequency or probability of the i-th character, $i = 1,2,3,4$;

d. $p_{.}$ = $p_{1.} + p_{2} + p_{3.} + p_{4.}$ =
= $p_{.1} + p_{.2} + p_{.3} + p_{.4}$ =
= 1.0 ;

e. $p_{i.}$ may be the same with $p_{j.}$ for $i \neq j$
$i = 1,2,3,4$ and $j = 1,2,3,4$, but it does not mean that $p_{ik} = p_{jk}$ for $k = 1,2,3,4$;

f. p_{ij} is the joint relative frequency or joint probability of the i-th character in the first position with the j-th character in the second position of 2-graph structure, $i = 1,2,3,4$ and $i = 1,2,3,4$

g. $\frac{p_{ij}}{p_{i\cdot}}$ is the relative frequency or probability of the j-th character in the second position of 2-graph structure, given the i-th character appears in the first position (i.e., the j-th character follows the given i-th character) and this is called **the row-wise conditional relative frequency or row-wise conditional probability** $i = 1,2,3,4$ and $j = 1,2,3,4$ (see Table 2.10 below) ;

h. $\frac{p_{ij}}{p_{\cdot j}}$ is the relative frequency or probability of the i-th character in the first position of 2-graph structure, given the j-th character appears in the second position (i.e., the i-th character is followed by the given j-th character) and this is called **the column-wise conditional relative frequency or column-wise conditional probability** $i = 1,2,3,4$ and $j = 1,2,3,4$ (see Table 2.11 below).

2.2 Some Measures from 1-graph and 2-graph Structures

Tables 2.1, 2.9, 2.10, and 2.11 can be used to provide two kinds of measurement, namely :

(a) statistical measures, and

(b) information theoretic measures.

2.2.1 Statistical measures

The relative frequency of every character from alphabet {A,B,C,D} in Table 2.1 can be used to give the characteristics of the characters in the alphabet. We shall consider the relative frequency as a measurement that can be applied to every character, and this variable will be included as a member in the measurement vector for analytical purposes. Let us call this variable as the subscripted variable **RELFREQ**[i], i=1,2,3,4.

Table 2.9 resembles the Two-way Classification design in experimental statistics [Broota, 1989]. The analysis of variance in this kind of experimental design can be divided into two parts, which are (a) Row-wise, and (b) Column-wise analyses of variance (see Tables 2.12 and 2.13). Tables 2.12 and 2.13 give some more characteristic measures of every character. Only the measures or variables, which have been shown analytically and empirically to be

Table 2.10: Row-wise Conditional Probability Space from 2-graph Structure

	1	2	3	4	Row Total
1	$\frac{p_{11}}{p_1}$	$\frac{p_{12}}{p_1}$	$\frac{p_{13}}{p_1}$	$\frac{p_{14}}{p_1}$	1.0
2	$\frac{p_{21}}{p_2}$	$\frac{p_{22}}{p_2}$	$\frac{p_{23}}{p_2}$	$\frac{p_{24}}{p_2}$	1.0
3	$\frac{p_{31}}{p_3}$	$\frac{p_{32}}{p_3}$	$\frac{p_{33}}{p_3}$	$\frac{p_{34}}{p_3}$	1.0
4	$\frac{p_{41}}{p_4}$	$\frac{p_{42}}{p_4}$	$\frac{p_{43}}{p_4}$	$\frac{p_{44}}{p_4}$	1.0

non-redundant, will be included in the measurement vector of every character. The names of the variables are :

(a) **ROWCOLSS**[i], stands for Row or Column Sum of Squares ;

(b) **WINROWSS**[i], the abbreviation of Within Row Sum of Squares ;

(c) **WINCOLSS**[i], the abbreviation of Within Column Sum of Squares;

The last two variables (b) and (c) are usually recognized as the effects of uncontrollable random errors for row and column, respectively.

Tables 2.10 and 2.11 also provide some measures which can be used as the characteristics of the alphabet characters. If for every character in Table 2.10 we add the square of each cell containing row-wise conditional probability, then we can obtain a measure or variable called the row-wise conditional probability sum of squares, **RWCPROSS**[i].

And if for every character in Table 2.11 we add the square of each cell containing column-wise conditional probability, then we can obtain a measure or variable called the column-wise conditional probability sum of squares, **CWCPROSS**[i].

Up to this point, we have six measures or variables which can be used to identify the alphabet characters.

Table 2.11: Column-wise Conditional Probability Space from 2-graph Structure

	1	2	3	4	Column Total
1	$\frac{p_{11}}{p_1}$	$\frac{p_{12}}{p_1}$	$\frac{p_{13}}{p_1}$	$\frac{p_{14}}{p_1}$	1.0
2	$\frac{p_{21}}{p_2}$	$\frac{p_{22}}{p_2}$	$\frac{p_{23}}{p_2}$	$\frac{p_{24}}{p_2}$	1.0
3	$\frac{p_{31}}{p_3}$	$\frac{p_{32}}{p_3}$	$\frac{p_{33}}{p_3}$	$\frac{p_{34}}{p_3}$	1.0
4	$\frac{p_{41}}{p_4}$	$\frac{p_{42}}{p_4}$	$\frac{p_{43}}{p_4}$	$\frac{p_{44}}{p_4}$	1.0

2.2.2 Information theoretic measures

If we take the negative logarithm of the relative frequency of every character in Table 2.1, then we will get a variable called **Information Content** [Jürgensen, *et al.*1984] [Shannon, 1948]. This is abbreviated by **INFOCONT**[i].

The entropy of a string of characters is defined as the summation of (relative frequency $\times$ information content) from every character in the alphabet of the string. In other words, every character will have its contribution to the entropy of the string. We can use this contribution as a characteristic, too. Let us call this variable contribution to entropy, **CTOENTRO**[i].

In Table 2.9,$\sum_{j=1}^{n}\sum_{i=1}^{m} p_{ij}(-\log p_{ij})$is called **the joint entropy** of the two-dimensional probability space.

$\sum_{j=1}^{n} p_{ij}(-\log p_{ij})$ is the contribution of the i-th row-element character to the joint entropy. The name of this variable is row element contribution to joint entropy, **RCTOJENT**[i], $i = 1,2,3,4$.

$\sum_{i=1}^{m} p_{ij}(-\log p_{ij})$ is the contribution of the j-th column-element character to the joint entropy. The name of this variable is column element contribution to joint entropy, **CC-**

Table 2.12: Row-wise Analysis of Variance from Two-dimensional Probability Space

	Source of Variation	degrees of freedom	Sum of squares
1	Rows	$(n-1)$	$\frac{1}{n}\sum_{i=1}^{n} p_{i.}^2 - \frac{G^2}{n^2}$ $= \sum_{i=1}^{n}(\frac{p_{i.}^2}{n} - \frac{G^2}{n^3})$
2	Within rows	$(n)(n-1)$	$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2 - \frac{1}{n}\sum_{i=1}^{n} p_{i.}^2$ $= \sum_{i=1}^{n}(\sum_{j=1}^{n} p_{ij}^2 - \frac{p_{i.}^2}{n})$
3	Total	n^2-1	$\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2 - \frac{G^2}{n^2}$ $= \sum_{i=1}^{n}(\sum_{j=1}^{n} p_{ij}^2 - \frac{G^2}{n^3})$
			Notes : G = Grand Total (in this case = 1.0) n = number of characters (in this case 4)

TOJENT[j], $j = 1, 2, 3, 4$.

In Table 2.10, **Conditional entropy** of all alphabet characters following a specific row character is defined as $\sum_{j=1}^{n} \frac{p_{ij}}{p_i}(-\log \frac{p_{ij}}{p_i})$.

The name of this variable is row-wise conditional entropy, abbreviated to **RWCENTRO[i]**, $i = 1, 2, 3, 4$.

In Table 2.11, **Conditional entropy** of all alphabet characters followed by a specific column character is defined as $\sum_{i=1}^{m} \frac{p_{ij}}{p_j}(-\log \frac{p_{ij}}{p_j})$.

The name of this variable is column-wise conditional entropy, **CWCENTRO[j]**, $j = 1, 2, 3, 4$.

By combining the statistical and information theoretic measures, we have obtained twelve variables to be included in the measurement vector of every character in the alphabet. It is expected that these variables will be able to identify the characters better than using only one variable, i.e., relative frequency of characters.

Table 2.13: Column-wise Analysis of Variance from Two-dimensional Probability Space

	Source of Variation	degrees of freedom	Sum of squares
1	Columns	$(n-1)$	$\frac{1}{n}\sum_{j=1}^{n} p_{.j}^2 - \frac{G^2}{n^2}$ $= \sum_{j=1}^{n}(\frac{p_{.j}^2}{n} - \frac{G^2}{n^3})$
2	Within cols	$(n)(n-1)$	$\sum_{j=1}^{n}\sum_{i=1}^{n} p_{ij}^2 - \frac{1}{n}\sum_{j=1}^{n} p_{.j}^2$ $= \sum_{j=1}^{n}(\sum_{i=1}^{n} p_{ij}^2 - \frac{p_{.j}^2}{n})$
3	Total	$n^2 - 1$	$\sum_{j=1}^{n}\sum_{i=1}^{n} p_{ij}^2 - \frac{G^2}{n^2}$ $= \sum_{j=1}^{n}(\sum_{i=1}^{n} p_{ij}^2 - \frac{G^2}{n^3})$
			Notes : G = Grand Total (in this case = 1.0) n = number of characters (in this case 4)

All of these formulas and information measures can contribute as the sources of variation in the process of discrimination and classification. In other words, these are random variables that can be parts or elements of feature vectors or measurement vectors in the process of identification of objects. In the next section we will see the application of these formulas to a simplified source text and crypto text. It can be shown that all of these formulas are invariant in both texts. This means that the formulas will give the same values before and after the process of transforming characters from source text into cryptotext.

2.3 Relationship between source text and crypto text

2.3.1 Monoalphabetic substitution

Suppose one has a string of eighty characters from alphabet { A, B, C, D } as follows :

'ABCDDCCBABCBBCAABCDCAAABCABBAACDDADDBBBC

BCADABCDAAAABBBDDDCCCACDBBDACACDACDABBBC'

The frequencies of characters A, B, C, and D are 22, 22, 20, and 16 respectively, and their relative frequencies are 0.275, 0.275, 0.250 and 0.200 respectively.

One can create a one-graph table of frequency and relative frequency for every character of the alphabet in the source text as shown in Table 2.14. Suppose that characters in the

Character	Frequency	Relative Frequency
A	22	0.2750
B	22	0.2750
C	20	0.2500
D	16	0.2000
Total	80	1.0000

Table 2.14: One-graph table of frequency and relative frequency of characters in the source text

string or source text are monoalphabetically substituted by the following non-singular rule

```
character A is substituted to character C ;
character B is substituted to character A ;
character C is substituted to character D ; and
character D is substituted to character B.
```

The substitution will result in the following string called crypto text :

'CADBBDDACADAADCCADBDCCCADCAACCDBBCBBAAAD

ADCBCADBCCCCAAABBBDDDCDBAABCDCDBCDBCAAAD' and the one-graph table of frequency and relative frequency for every character of the alphabet in the crypto text is shown in Table 2.15. By analyzing the two tables using only the frequency or relative frequency for every character, one can identify that character B in the crypto text table is the substitution of character D in the source text.

Character	Frequency	Relative Frequency
A	22	0.2750
B	16	0.2000
C	22	0.2750
D	20	0.2500
Total	80	1.0000

Table 2.15: One-graph table of frequency and relative frequency of characters in the crypto text

Similarly, character D in the crypto text is the substitution of character C in the source text. But, character A and C in the crypto text have the same frequencies or relative frequencies. This situation shows the difficulty of identifying characters by using only one variable in this case frequency or relative frequency – when there is not enough variation generated by the variable.

To obtain more variation which implies more information – one should look for dependency or relationship between characters both in source and crypto texts. This can be accomplished by generating more variables using digraph frequency or relative frequency such as given in Table 2.16.

The corresponding digraphs for relative frequencies are given in Table 2.17 and Table 2.18. It is observable that the Row Total as well as the Column Total of frequencies and relative frequencies are the same with the previous ones in the one-graph tables, namely Table 2.14 and Table 2.15.

Using the table one can generate the following random variables or sources of variation. The definitions of these sources of variation were given in the previous section (Section 2.2):

01. RELFREQ   02. ROWCOLSS   03. WINROWSS   04. WINCOLSS

SOURCE TEXT						CRYPTO TEXT					
ABCDDCCBABCBBCAABCDC AAABCABBAACDDADDBBBC BCADABCDAAAABBBDDDCC CACDBBDACACDACDABBBC						CADBBDDACADAADCCADBD CCCADCAACCDBBCBBAAAD ADCBCADBCCCCAAABBBDD DCDBAABCDCDBCDBCAAAD					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	7	8	5	2	22	A	9	2	2	9	22
B	2	9	9	2	22	B	2	5	6	3	16
C	7	3	3	7	20	C	8	2	7	5	22
D	6	2	3	5	16	D	3	7	7	3	20
Column Total	22	22	20	16	80	Column Total	22	16	22	20	80

Table 2.16: Digraph frequency generated from source text and crypto text

05. RWCPROSS	06. CWCPROSS	07. INFOCONT	08. CTOENTRO
09. RCTOJENT	10. CCTOJENT	11. RWCENTRO	12. CWCENTRO

The values of these variables for a particular character are then arranged in a vector called *measurement vector* for that character.

It can be shown that the measurement vectors can be used for identifying characters in the crypto text based on the information from the source text. The variables in the measurement vectors have an important **invariant** feature, that is, each of them preserves the pattern of values for both substituted and substituting characters. In other words, it is possible to recognize the pattern for every character in crypto text based on the pattern for every character in source text by manipulating the values of the variables.

Since every character in the alphabet has more than one random variable, it is possible to identify the characters in the crypto text by using a statistical procedure called multivariate analysis based on the samples of source text from the same population. One of the multivariate statistical procedures which is appropriate to explore the differences of characters in the alphabet is called discriminant analysis. Furthermore, a technique called

	A	B	C	D	Row total
A	0.0875	0.1000	0.0625	0.0250	0.2750
B	0.0250	0.1125	0.1125	0.0250	0.2750
C	0.0875	0.0375	0.0375	0.0875	0.2500
D	0.0750	0.0250	0.0375	0.0625	0.2000
Column Total	0.2750	0.2750	0.2500	0.2000	1.0000

Table 2.17: Digraph relative frequency generated from source text

	A	B	C	D	Row total
A	0.1125	0.0250	0.0250	0.1125	0.2750
B	0.0250	0.0625	0.0750	0.0375	0.2000
C	0.1000	0.0250	0.0875	0.0625	0.2750
D	0.0375	0.0875	0.0875	0.0375	0.2500
Column Total	0.2750	0.2000	0.2750	0.2500	1.0000

Table 2.18: Digraph relative frequency generated from crypto text

Stepwise Linear Discriminant Analysis is helpful to select the random variables which have significant contribution in distinguishing the characters of the alphabet. The discriminant analysis will become the topic of discussion in the following chapter.

In the example discussed so far, the alphabet consists only four characters A, B, C, and D. It could be extended to include all characters that can be used in some natural language, such as English, text. In this research, the alphabet is extended to consist of twenty seven characters A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z, and Blank or Space.

Figure 2.1 shows the **structural signature of text** based on two-dimensional probability space generated by 2-graph structure of texts, along with the **measurement vector** which consists of twelve variables. Figure 2.2 presents the sample distribution of the twelve variables of only five characters or letters which will involve in linear discriminant analysis. The size of the sample of texts is two hundred. We can roughly say, based on the sample distribution, that the individual distribution of every variable follows a symmetric, bell-shaped distribution or the normal distribution, which is very important for further analysis.

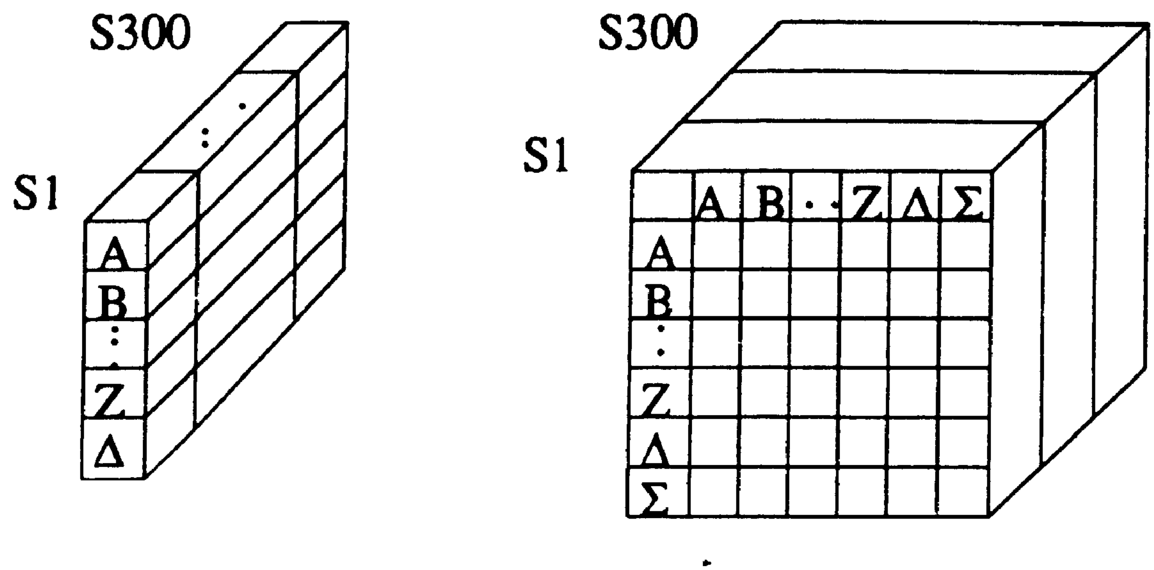


Figure 2.1: Structural Signature of Text

Measurement Vector consists of the variables :

Relative Frequency (RELFREQ)

Row or Column Sum of Squares (ROWCOLSS)

Within Row Sum of Squares (WINROWSS)

Within Column Sum of Squares (WINCOLSS)

Row-wise Conditional Probability Sum of Squares (RWCPROSS)

Column-wise Conditional Probability Sum of Squares (CWCPROSS)

Information Content (INFOCONT)

Contribution to Entropy (CTOENTRO)

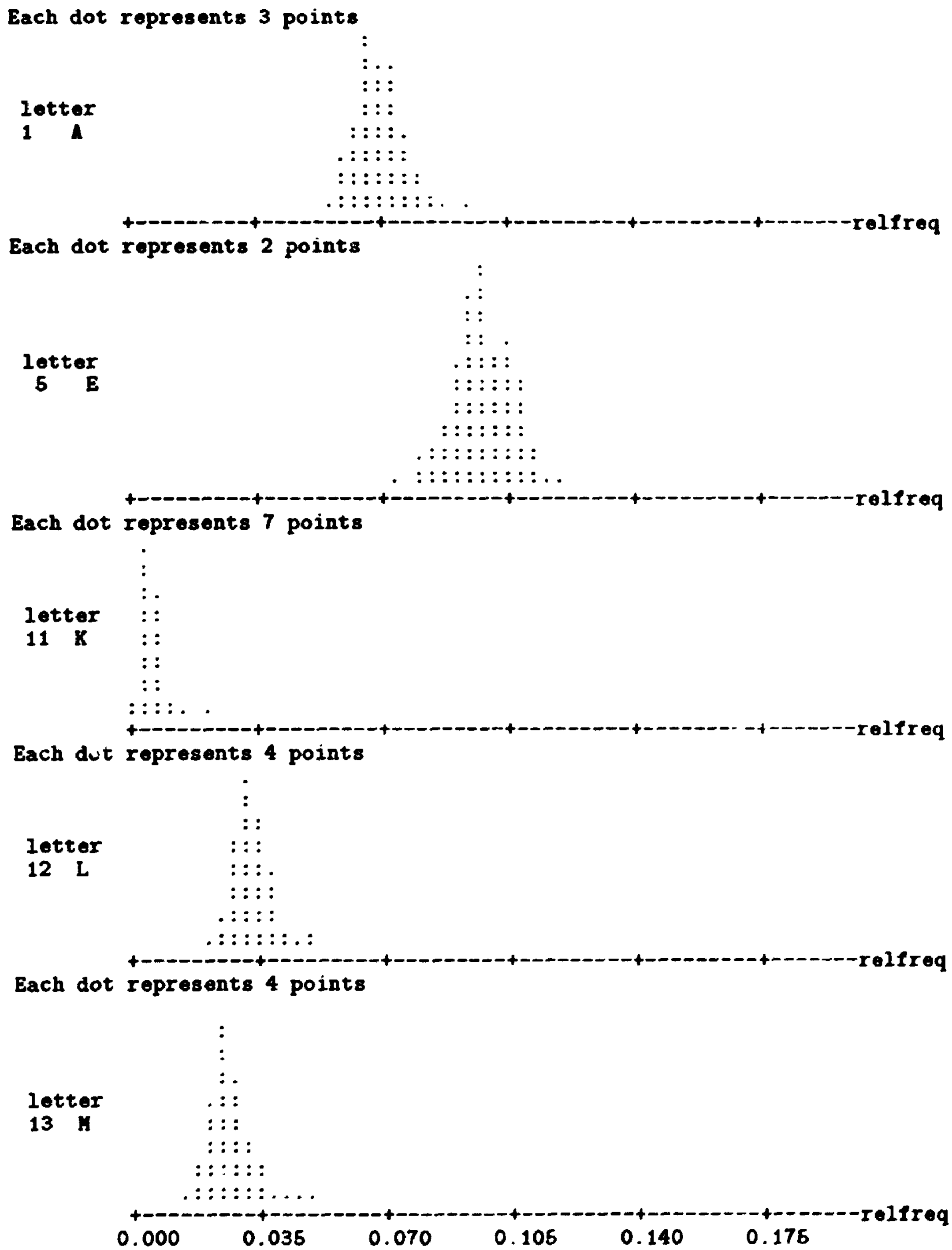
Row Contribution to Joint Entropy (RCTOJENT)

Column Contribution to Joint Entropy (CCTOJENT)

Row-wise Conditional Entropy (RWCENTRO)

Column-wise Conditional Entropy (CWCENTRO)

Figure 2.2: Sample distribution of 12 variables of 5 characters involved in liniear discriminant analysis



(continued)

```
Each dot represents 4 points
                .
                :
                ::
 letter        :::
 1   A         :::
               ::::
              :::::.
              ::::::.   .
         +---------+---------+---------+---------+---------+-------rowcolss
Each dot represents 2 points
                       :
                      .:
                      ::
                      :::
 letter               ::::
 5  E                .:::::
                    .::::::
                    ::::::::
                   .:::::::::.
                 ..:::::::::: .
         +---------+---------+---------+---------+---------+-------rowcolss
Each dot represents 14 points
         .
         :
 letter  :
 11  K   :
         :
         :
         :
         :.
         +---------+---------+---------+---------+---------+-------rowcolss
Each dot represents 7 points
           .
          .:
          ::
 letter   ::
 12  L    ::
          ::
          ::
          :::.
         +---------+---------+---------+---------+---------+-------rowcolss
Each dot represents 11 points
          .
          :
 letter   :
 13  M    :
          :
          :
          :.
         .::..
         +---------+---------+---------+---------+---------+-------rowcolss
        0.00000   0.00025   0.00050   0.00075   0.00100   0.00125
```

(continued)

Each dot represents 2 points

```
                    .
                    :
                   .:
                   :: .
  letter           ::::
  1   A            :::: .
                  :::::::
                  :::::::
                 ::::::::..
               .:::::::::::::: .       .
          +---------+---------+---------+---------+---------+-------winrowss
```

```
                                     .
                                     : .
                                     : :
  letter                         :  .:.:
   5  E                        .::..:::. .
                               ::::::::: ::: ::
                               ::::::::::::: :::   :
                          ...::::::::::::::::.:::..::
                        ..::::::::::::::::::::::::::.:.    .
          +---------+---------+---------+---------+---------+-------winrowss
```

Each dot represents 12 points

```
          .
          :
  letter  :
  11  K   :
          :
          :
          :
          ::.
          +---------+---------+---------+---------+---------+-------winrowss
```

Each dot represents 5 points

```
             :
             :
            .:
  letter    ::
  12  L     :::
            :::
            ::.
           :::::::.. .
          +---------+---------+---------+---------+---------+-------winrowss
```

Each dot represents 6 points

```
            .
            :
            :
  letter    :
  13  M     :
            :.
           :::
           :::.
           ::::.: ..   .
          +---------+---------+---------+---------+---------+-------winrowss
          0.00000   0.00035   0.00070   0.00105   0.00140   0.00175
```

(continued)

```
Each dot represents 2 points
                   :
                   :
                 :.: :
                 :::.:
 letter          ::::::
 1   A           ::::::.
                .:::::::
                ::::::::..
              .:::::::::::.:
         +---------+---------+---------+---------+---------+---------+-------wincolss
Each dot represents 2 points
                   .
                   :
                   :
                 .::
 letter          :::
  5  E           :::  .
                .::::
                ::::::.
                ::::::::
                ::::::::.
              .::::::::::.:.
         +---------+---------+---------+---------+---------+---------+-------wincolss
Each dot represents 13 points
         .
         :
         :
 letter  :
 11  K   :
         :
         :
         ::.  .
         +---------+---------+---------+---------+---------+---------+-------wincolss
Each dot represents 6 points

           :
           :
 letter    :
 12  L     :.
           ::
          .::
          ::::
          :::::..
         +---------+---------+---------+---------+---------+---------+-------wincolss
Each dot represents 5 points
           .
           :
           :
 letter    ::
 13  M     ::
          ::::
          ::::
          ::::::: .. .            .
         +---------+---------+---------+---------+---------+---------+-------wincolss
        0.00000   0.00040   0.00080   0.00120   0.00160   0.00200
```

(continued)

```
Each dot represents 6 points
                :
               .:
               ::
 letter        ::
 1   A         ::
               ::.
               :::
              .:::..
          +---------+---------+---------+---------+---------+-------rwcpross
Each dot represents 5 points
                 .
                 :
                ::
 letter         :::
 5   E          :::
                :::
                :::.
               :::::. .
          +---------+---------+---------+---------+---------+-------rwcpross
                         .
                         :
                        ::
                      : :.
                      : ::
 letter             . : ::: ..
 11  K              :.:.::: ::
                    ::::::: :: :
                    ::::::::::.:
                   :::::::::::::
                                          .
                  .:::::::::::::.  :  : :
                  :::::::::::::::::::::.:  .  ..  .          .
          +---------+---------+---------+---------+---------+-------rwcpross
Each dot represents 5 points
                :
                :.
                ::
 letter         ::
 12  L          ::.
                :::
                :::.
               :::::..
          +---------+---------+---------+---------+---------+-------rwcpross
Each dot represents 4 points
                 ..
                 ::
                 ::
 letter         .::
 13  M          :::
                ::::
                ::::.
               .::::::..  .
          +---------+---------+---------+---------+---------+-------rwcpross
         0.00      0.20      0.40      0.60      0.80      1.00
```

(continued)

```
Each dot represents 6 points
               .
               :
               :
 letter        :
 1   A        :::
              :::
              ::::
             .::::.
         +---------+---------+---------+---------+---------+-------cwcpross
Each dot represents 11 points
             :
             :
             :
 letter      :
  5  E       :
             :
             :
            :::
         +---------+---------+---------+---------+---------+-------cwcpross
Each dot represents 4 points
                   : :
                   : :
                   : :
                   : :
                 . : :
 letter          : : : .
 11  K           :.:::.:    .
                 ::::::::..::
                 ::::::::::::
                 ::::::::::::
                .:::::::::::: ::. :... :
                :::::::::::::.:::::::::::.  . :        .
         +---------+---------+---------+---------+---------+-------cwcpross
Each dot represents 5 points
               .
               :
               :
 letter        :
 12  L         ::
              :::
              :::
              ::::
             .::::...  .
         +---------+---------+---------+---------+---------+-------cwcpross
Each dot represents 3 points
                   .
                   :
                 :::
 letter          :::.
 13  M           ::::.
                ::::::
                :::::::
               .::::::::.:..    . .
         +---------+---------+---------+---------+---------+-------cwcpross
        0.00      0.20      0.40      0.60      0.80      1.00
```

(continued)

```
Each dot represents 7 points
                          .
                          :
                         .:
 letter                  ::
 1   A                   ::
                         ::
                         ::
                        .:::
         +---------+---------+---------+---------+---------+-------infocont
Each dot represents 6 points
                       .
                       ::
                       ::
 letter                ::
  5  E                 ::
                       ::
                       ::
                       ::
                      .:::
         +---------+---------+---------+---------+---------+-------infocont
Each dot represents 2 points
                                           .
                                           :
                                           :.
 letter                                 :.:::
 11  K                                 ::::::
                                       ::::::.
                                      .:::::::.
                                      :::::::::
                                 .  ..:::::::::::..  .
         +---------+---------+---------+---------+---------+-------infocont
Each dot represents 5 points

                              .
                              ::
                              ::
 letter                       ::
 12  L                        ::
                             .::
                             ::::
                            :::::.
         +---------+---------+---------+---------+---------+-------infocont
Each dot represents 5 points
                                :
                                :
                               .:
 letter                        :::
 13  M                         :::
                              :::::
                            .:::::::..
         +---------+---------+---------+---------+---------+-------infocont
        0.0       1.6       3.2       4.8       6.4       8.0
```

(continued)

Each dot represents 3 points

letter
1 A

```
                              .
                              ::.
                              :::
                              :::
                             .::::
                            .:::::
                            :::::::
                           .::::::::...
+---------+---------+---------+---------+---------+-------ctoentro
```

Each dot represents 3 points

letter
5 E

```
                                      .
                                      :
                                     ::
                                     ::.
                                     ::::
                                    .::::
                                    :::::
                                   ::::::.
                                .  :::::::::.
+---------+---------+---------+---------+---------+-------ctoentro
```

Each dot represents 3 points

letter
11 K

```
   .:
   ::.
   :::.
   ::::
   ::::
   ::::.
  ::::::
 ::::::::..  .
+---------+---------+---------+---------+---------+-------ctoentro
```

Each dot represents 3 points

letter
12 L

```
                  :
                  :::
                 ::::
                 ::::.
                .:::::
               .::::::
             ..:::::::::.:.
+---------+---------+---------+---------+---------+-------ctoentro
```

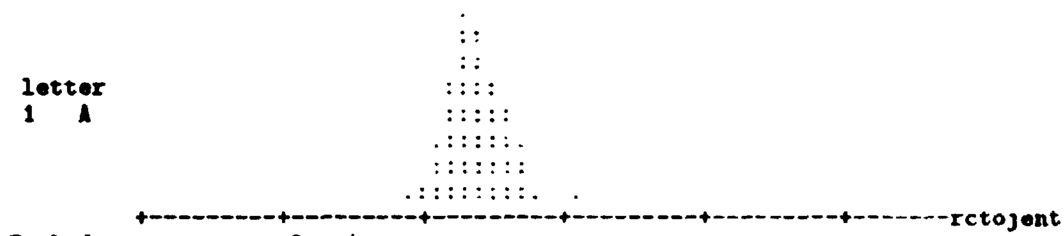
Each dot represents 3 points

letter
13 M

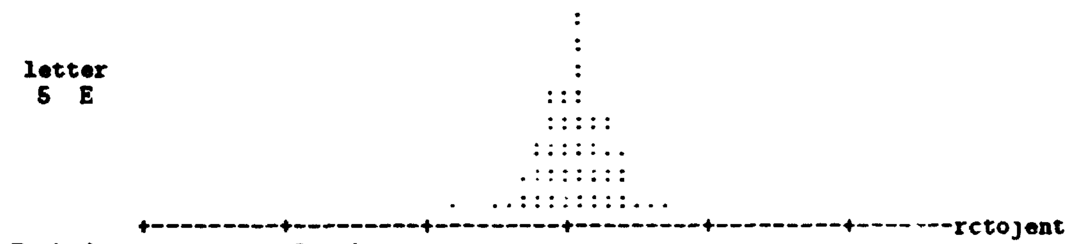
```
               .
               :
               :
               :
               ::.
              ::::
             ::::::
            ::::::::
          ..:::::::::::.. ..
+---------+---------+---------+---------+---------+-------ctoentro
0.000     0.060     0.120     0.180     0.240     0.300
```

(continued)

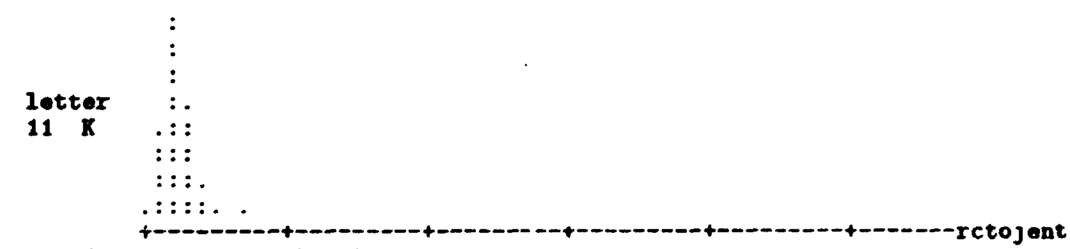
Each dot represents 3 points



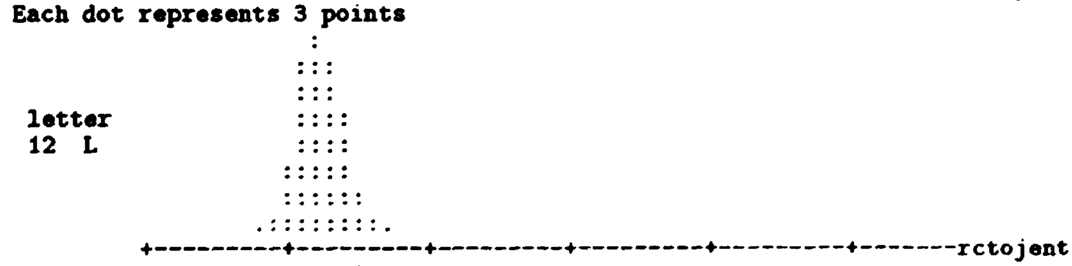
Each dot represents 3 points



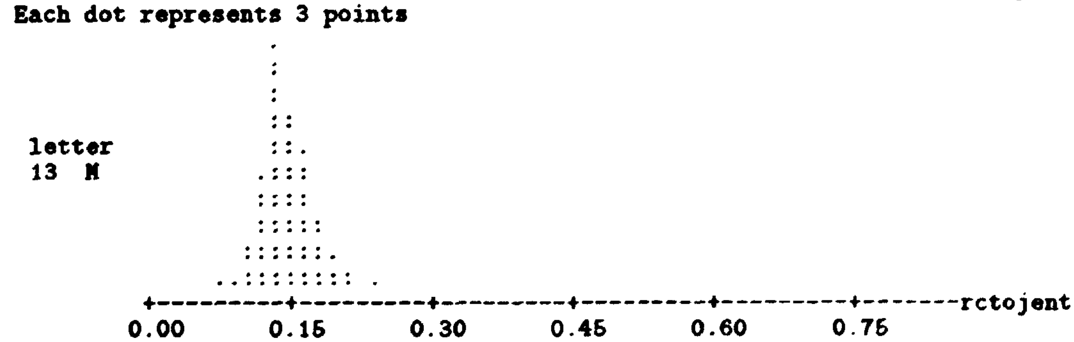
Each dot represents 6 points



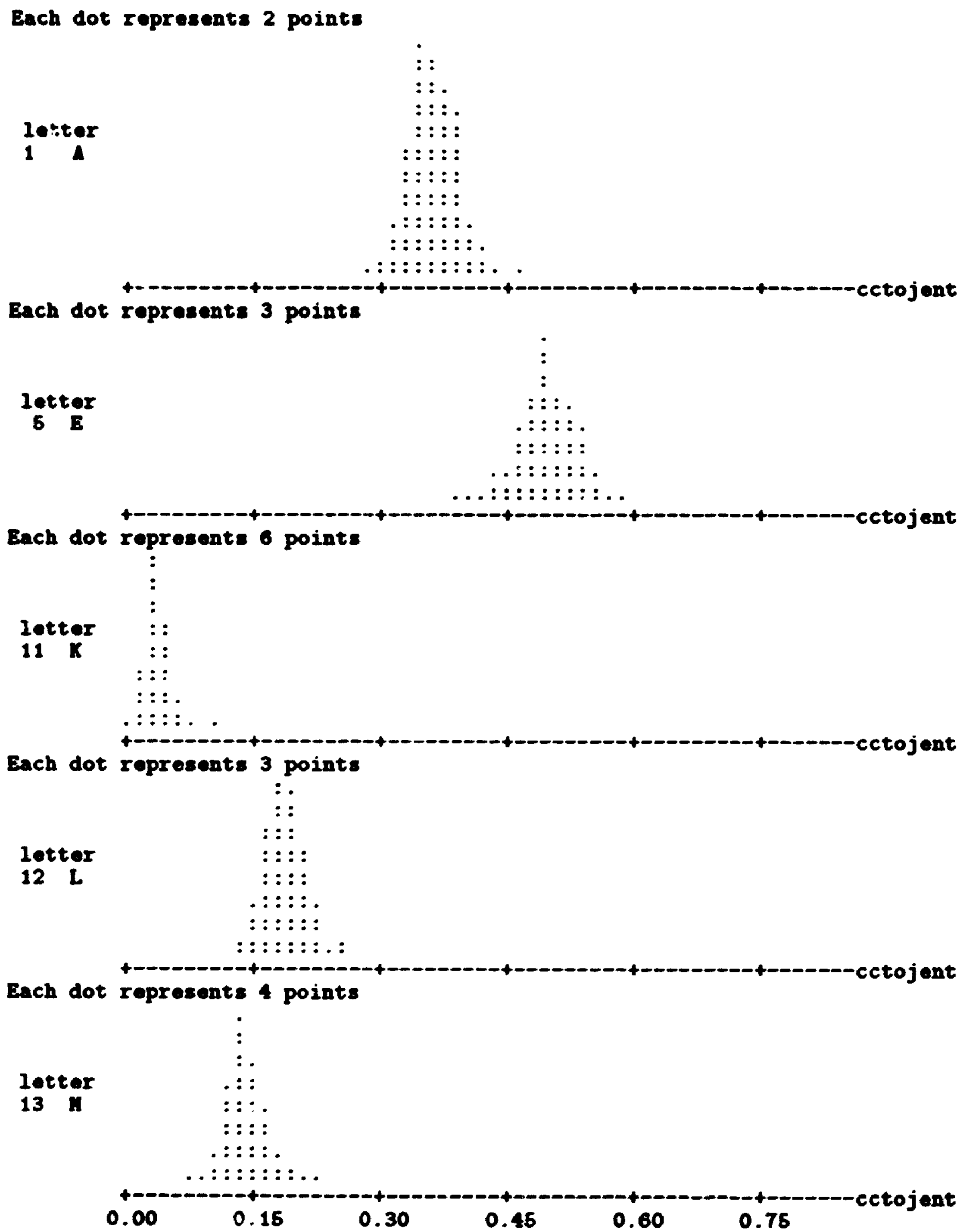
Each dot represents 3 points



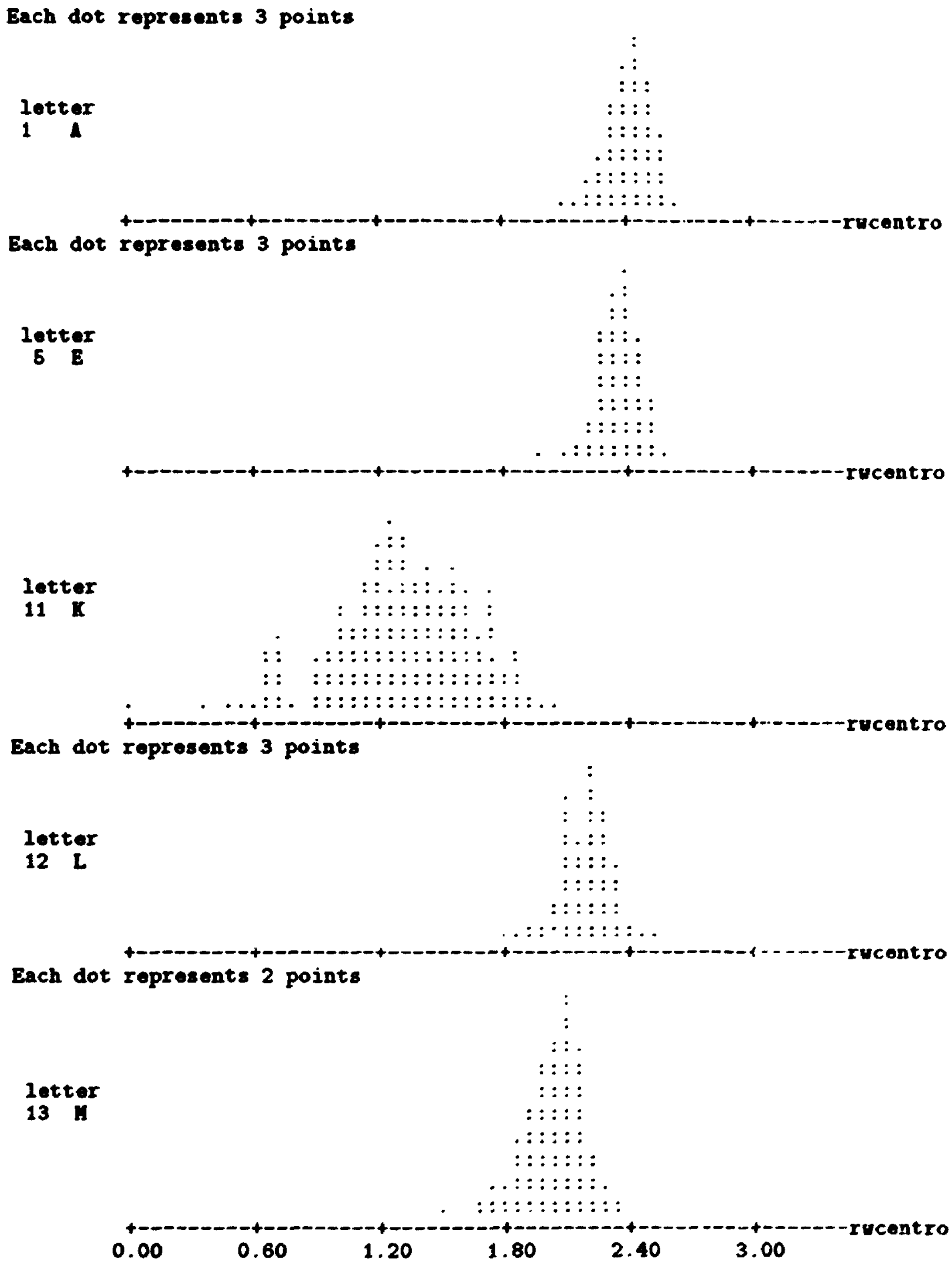
Each dot represents 3 points



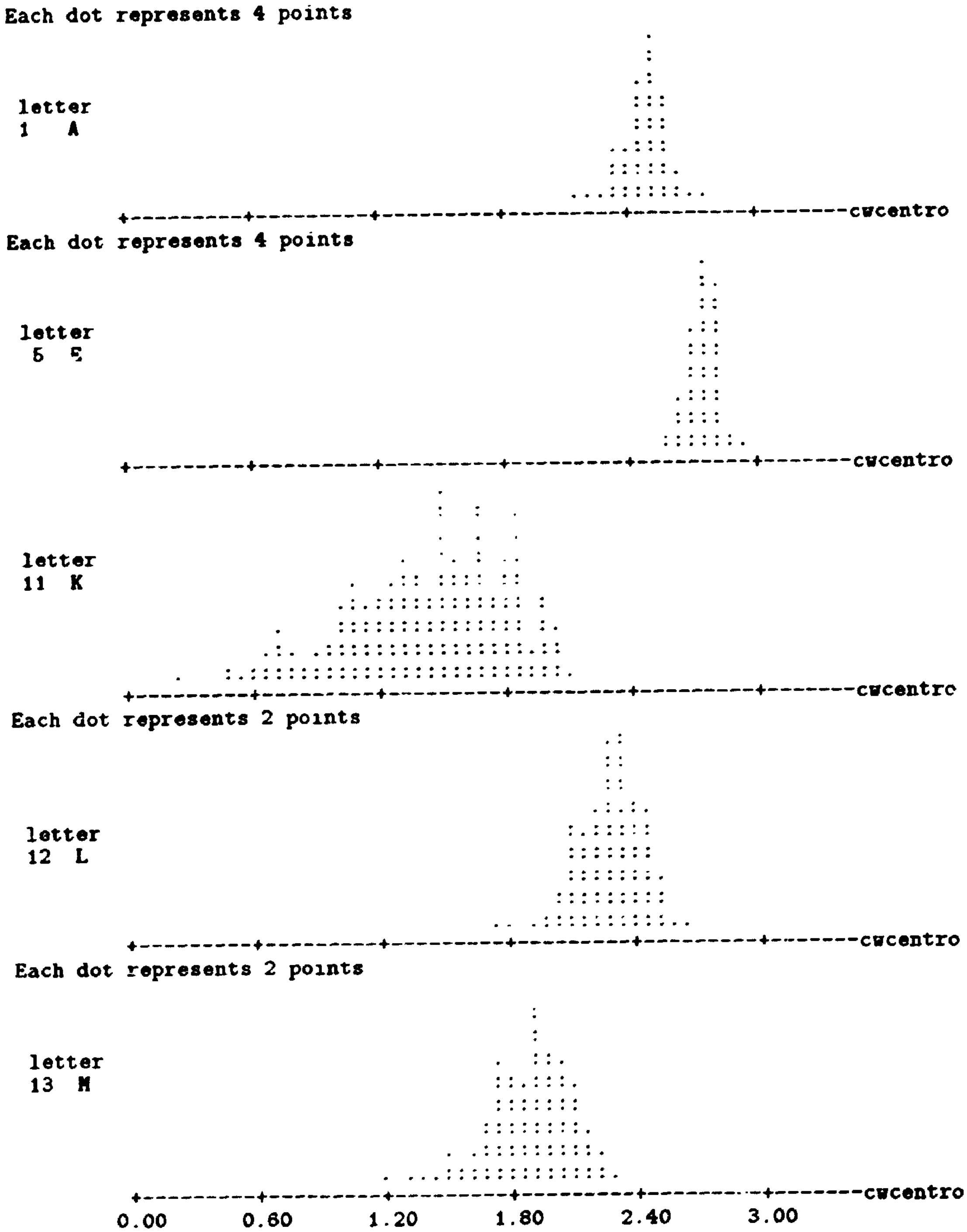
(continued)


Each dot represents 2 points
letter
1 A
cctojent
Each dot represents 3 points
letter
5 E
cctojent
Each dot represents 6 points
letter
11 K
cctojent
Each dot represents 3 points
letter
12 L
cctojent
Each dot represents 4 points
letter
13 M
cctojent
0.00
0.15
0.30
0.45
0.60
0.75

(continued)


Each dot represents 3 points
letter
1 A
rwcentro
Each dot represents 3 points
letter
5 E
rwcentro
letter
11 K
rwcentro
Each dot represents 3 points
letter
12 L
rwcentro
Each dot represents 2 points
letter
13 M
rwcentro
0.00
0.60
1.20
1.80
2.40
3.00

(continued)


Each dot represents 4 points
letter
1 A
cwcentro
Each dot represents 4 points
letter
5 E
cwcentro
letter
11 K
cwcentro
Each dot represents 2 points
letter
12 L
cwcentro
Each dot represents 2 points
letter
13 M
cwcentro
0.00
0.60
1.20
1.80
2.40
3.00

2.3.2 Polyalphabetic Substitution

Let the source text is the same as before. Suppose the source text is substituted iteratively by using three different rules in sequential fashion (key length = 3).

SOURCE TEXT :

```
ABCDDCCBABCBBCAABCDCAAABCABBAACDDADDBBBC
1231231231231231231231231231231231231231
BCADABCDAAAABBBDDDCCCACDBBDACACDACDABBBC
2312312312312312312312312312312312312312
```

Let the substitution rules are as the following :

```
For position 1 :  A becomes B ;
                  B becomes A ;
                  C becomes D ;
                  D becomes C ;
For position 2 :  A becomes C ;
                  B becomes A ;
                  C becomes D ;
                  D becomes B ;
For position 3 :  A becomes B ;
                  B becomes C ;
                  C becomes D ;
                  D becomes A .
```

The result of the polyalphabetic substitution is the following crypto text .

```
BADCBDDABADCADBBADCDBBCCDCCACBDBABBAAACD
1231231231231231231231231231231231231231

ADBBBADABCBBACABACDDDCDCACCCDBDABDABACAD
2312312312312312312312312312312312312312
```

Since 80 is not a multiple of 3, the last two characters in the source and crypto texts will be removed from the analysis of the relationship between the texts.

Information about the relationship between the texts is obtained by dividing both of the texts into six that is two times key length - parts and then by creating a digraph table of character frequency or relative frequency for every part such as given in Table 2.19 up to Table 2.24.

The first, the second, and the third parts represent the self-dependencies for positions 1, 2, and 3 respectively. The fourth part represents the dependency between position 1 and position 2 ; the fifth one represents the dependency between positions 2 and 3 ; and the sixth one between positions 3 and 1.

Some patterns of relationship between the source and crypto texts can be observed from the tables. These patterns will be useful in determining the generation of random variables for characters in every position, that is positions 1, 2, and 3. The row and column elements of the digraph in Part 1 together with the row elements of the digraph in Part 4 and the column elements of the digraph in Part 6 will determine the generation of random variables for identification of characters in position 1. Similarly, the row and column elements of the digraph in Part 2 along with the row elements of the digraph in Part 5 and the column elements of the digraph in Part 4 will determine the generation of random variables for identification of characters in position 2. The row and column elements of the digraph in Part 3 along with the row elements of the digraph in Part 6 and the column elements of the digraph in Part 5 determine the generation of random variables for identification of characters in position 3.

The number of variables that can be generated from Parts 1, 2, and 3 is the same as the number of variables generated in the monoalphabetic substitution cipher, that is twelve variables. The additional variables are generated from the row elements and column elements of the other parts. As an example, for identifying the characters in position 1, twelve variables can be generated from the digraph relative frequency of Part 1, four variables can be generated from the row elements of the digraph relative frequency of Part 4, and four more variables from the column elements of the digraph relative frequency of

PART 1 :

SOURCE TEXT :	CRYPTO TEXT :
ADCBBADACBCABCABAABDCDDAAA	BCDAABCBDADBADBABBACDCCBBB
11111111111111111111111111	11111111111111111111111111

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	4	3	1	2	10	A	1	2	1	2	6
B	2	1	2	1	6	B	3	4	2	1	10
C	2	2	0	1	5	C	0	2	1	2	5
D	2	0	2	1	5	D	2	2	1	0	5
Column Total	10	6	5	5	26	Column Total	6	10	5	5	26

Table 2.19: Digraph of frequency representing self-dependency for position 1 in source text and crypto text

PART 2 :

SOURCE TEXT :	CRYPTO TEXT :
BDBCCBCAAADDBBDCABDCABACCB	ABADDADCCCBBAABDCABDCACDDA
22222222222222222222222222	22222222222222222222222222

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	2	2	1	1	6	A	2	3	1	2	8
B	1	2	2	3	8	B	2	1	0	2	5
C	3	2	2	0	7	C	2	1	2	1	6
D	0	2	2	1	5	D	2	0	3	2	7
Column Total	6	8	7	5	26	Column Total	8	5	6	7	26

Table 2.20: Digraph of frequency representing self-dependency for position 2 in source text and crypto text

PART 3 : SOURCE TEXT : CCABACABBADDBCADABDCCBCDDB 33333333333333333333333333 CRYPTO TEXT : DDBCBDBCCBAACDBABCADDCDAAC 33333333333333333333333333

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	0	3	1	2	6	A	2	1	2	1	6
B	2	1	3	1	7	B	2	0	3	1	6
C	3	1	2	1	7	C	1	2	1	3	7
D	1	2	1	2	6	D	1	3	1	2	7
Column Total	6	7	7	6	26	Column Total	6	6	7	7	26

Table 2.21: Digraph of frequency representing self-dependency for position 3 in source text and crypto text

PART 4 : SOURCE TEXT : ABDDCBBCBCABDCAACABACDADBB 12121212121212121212121212 CBADBCAAABBDDCCADBDAACACAB 12121212121212121212121212 CRYPTO TEXT : BACBDAADADBACDBCDCACDBBBAA 12121212121212121212121212 DABBADBCBAABCDDCCACCBDBDBA 12121212121212121212121212

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	2	4	2	2	10	A	1	1	1	3	6
B	1	1	3	1	6	B	4	2	2	2	10
C	2	2	0	1	5	C	1	1	1	2	5
D	1	1	2	1	5	D	2	1	2	0	5
Column Total	6	8	7	5	26	Column Total	8	5	6	7	26

Table 2.22: Digraph of frequency representing dependency between positions 1 and 2 for source text and crypto text

PART 5 :

SOURCE TEXT :
BCDCBACBCABCCAABABAADDDDBB
23232323232323232323232323
BCDACDAABBDDCCACBBACCDCDBB
23232323232323232323232323

CRYPTO TEXT :
ADBDABDCDBADDBCCCCCBBABAAC
23232323232323232323232323
ADBBDACBACBADDCDACCDDADAAC
23232323232323232323232323

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	2	2	2	0	6	A	0	1	4	3	8
B	1	4	3	0	8	B	3	1	0	1	5
C	2	1	1	3	7	C	0	2	2	2	6
D	1	0	1	3	5	D	3	2	1	1	7
Column Total	6	7	7	6	26	Column Total	6	6	7	7	26

Table 2.23: Digraph of frequency representing dependency between positions 2 and 3 for source text and crypto text

PART 6 :

SOURCE TEXT :
CDCCABBBAACDAABCBBACDADBBC
31313131313131313131313131
CAABDAAABBDDCCCDBDCADADABA
31313131313131313131313131

CRYPTO TEXT :
DCDDBACABBDCBBCDCABDABAACD
31313131313131313131313131
DBBAABBBCAACDDDCCCDBABABCB
31313131313131313131313131

SOURCE TEXT						CRYPTO TEXT					
	A	B	C	D	Row Total		A	B	C	D	Row Total
A	3	2	1	0	6	A	1	4	1	0	6
B	1	3	2	1	7	B	2	3	0	1	6
C	2	0	2	3	7	C	3	1	1	2	7
D	4	1	0	1	6	D	0	2	3	2	7
Column Total	10	6	5	5	26	Column Total	6	10	5	5	26

Table 2.24: Digraph of frequency representing dependency between positions 3 and 1 for source text and crypto text

Part 6.

The names and the definition of the twelve variables are the same as the ones given in the monoalphabetic substitution cipher, but with a slight modification in the names :

1. RELFREQ	2. ROWCOLSS	3. WINROWS3	4. WINCOLS3
5. RWCPROS3	6. CWCPROS3	7. INFOCONT	8. CTOENTRO
9. RCTOJEN3	10. CCTOJEN3	11. RWCENTR3	12. CWCENTR3

The names of four variables generated from the row elements are :

1. WINROWSS	2. RWCPROSS	3. RCTOJENT	4. RWCENTRO

and the names of four variables generated from the column elements are :

1. WINCOLSS	2. CWCPROSS	3. CCTOJENT	4. CWCENTRO

It can also be shown that the measurement vectors of these variables are **invariant** between character groups in source text and crypto text at every position.

CHAPTER 3

DISCRIMINANT ANALYSIS AND CLASSIFICATION

3.1 Introduction

This chapter gives an explanation about discriminant analysis and classification from theoretical point of view. The explanation is for two groups or populations, as the starting point. Methods involving more than two populations — such as this research, which involves twenty seven character groups — will be discussed in the next chapter.

Discriminant analysis and classification are multivariate statistical techniques concerned with *separating* distinct sets of objects (or observations) and with *allocating* new objects (observations) to previously defined groups. [Johnson *et al*, 1988]

In multivariate statistical methods, we define p-dimensional random variable $\mathbf{X}$ as the vector

$$\mathbf{X}' = [X_1, X_2, \cdots, X_p]$$

whose elements are continuous unidimensional random variables.

In this research, the p-dimensional random variable $\mathbf{X}$ is the measurement vector consisting of the sources of variation mentioned in Chapter 2, and p has different values such as the following :

(a) $p = 1$ for both monoalphabetic and polyalphabetic substitutions : $\mathbf{X}' = [X_1]$ where X_1 is RELFREQ or relative frequency of every character group ;

(b) $p = 3$ for both substitutions : $\mathbf{X}' = [X_1, X_2, X_3]$ where X_1 = RELFREQ, X_2 = INFOCONT, and X_3 = CTOENTRO ;

(c) $p = 12$ for monoalphabetic substitution : $\mathbf{X}' = [X_1, X_2, \cdots, X_{12}]$ and the name of variables $X_1, X_2, \cdots, X_{12}$ were given in Chapter 2.

(d) $p = 20$ for polyalphabetic substitution : $\mathbf{X}' = [X_1, X_2, \cdots, X_{20}]$ (given in Chapter 2).

The parameter *expected value* of the random vector $\mathbf{X}$ is merely the vector of the expectations of its elements :

$$E(\mathbf{X}') = [E(X_1), E(X_2), \cdots, E(X_p)]$$

The parameter *covariance* of the elements X_i and X_j of $\mathbf{X}$ is defined as

$$\begin{aligned}
cov(X_i, X_j) &= E\{[X_i - E(X_i)][X_j - E(X_j)]\} \\
&= E(X_i X_j) - [E(X_i)][E(X_j)] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_i x_j f_{ij}(x_i, x_j) dx_i\, dx_j - [E(X_i)][E(X_j)] \\
&= \sigma_{ij}
\end{aligned}$$

where $f_{ij}(x_i, x_j)$ is the joint density of X_i and X_j. If $i = j$, the covariance is the *variance* of X_i, and we shall customarily write $\sigma_{ii} = \sigma_i^2$. The extension of the variance notion to the p-component random vector $\mathbf{X}$ is the matrix of variances and covariances

$$\begin{aligned}
E\{[\mathbf{X} - E(\mathbf{X})][\mathbf{X} - E(\mathbf{X})]'\} &= \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \cdots & \sigma_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix} \\
&= \mathbf{\Sigma}
\end{aligned}$$

We shall call this symmetric matrix the *covariance matrix* of $\mathbf{X}$.

The variance of a linear compound or linear combination $\ell'\mathbf{X} = \ell_1 X_1 + \ell_2 X_2 + \cdots + \ell_p X_p$ of random variables is :

$$\begin{aligned}
var(\ell'\mathbf{X}) &= \sum_{i=1}^{p}\sum_{j=1}^{p} \ell_i \ell_j \sigma_{ij} \\
&= \ell'\mathbf{\Sigma}\ell
\end{aligned}$$

[Morrison, 1967].

3.2 General Ideas

Suppose we have g populations or groups, $\pi_1, \pi_2, \cdots, \pi_g$ ($g = 27$ in this research). Lachenbruch [1975] outlined the analysis as follows (see also the model diagram in Chapter 1, Figure 1.2) :

(a) One observes a $p \times 1$ vector of random variables $\mathbf{X}' = [X_1, X_2, \cdots, X_p]$ and must assign the individual whose measurements are given by $\mathbf{x}$ to one of the populations $\pi_i, i = 1, 2, \cdots, g$.

(b) One needs a rule to classify $\mathbf{x}$ to one of the populations $\pi_i, i = 1, 2, \cdots, g$.

1. If the parameters of the distribution of $\mathbf{x}$ in $\pi_1, \pi_2, \cdots, \pi_g$ are known, one may use this knowledge in the construction of a classification rule ;
2. If the parameters are not known, one uses samples of size n_i from π_i to estimate the parameters.

(c) One needs a criterion of goodness of classification.

1. Fisher [1936] suggested using a linear combination of variables of the observations, and choosing the coefficients so that the ratio of the difference of the means of the linear combination in the g groups to its variance is maximized. In other words, Fisher's idea was to transform the multivariate observation $\mathbf{x}$ to univariate observation y such that the y's derived from populations $\pi_1, \pi_2, \cdots, \pi_i$ were separated as much as possible. In Fisher's approach, the linear combination is denoted by a new random variable $Y = \ell'\mathbf{X}$.
2. Welch [1939] suggested that minimizing the total probability of misclassification would be a sensible idea. Von Mises [1945] suggested minimizing the maximum probability of misclassification in the two groups. Various authors have suggested that different types of misclassification have different costs and advocated minimizing the total cost of misclassification. A clear discussion of this and the use of the Bayes theorem approach were given by Anderson [1958]. In the Bayes

theorem approach, the classification rule is to assign $\mathbf{x}$ to the group with the largest posterior probability. By definition, the conditional density of $\mathbf{x}$ given π_i is $f_i(\mathbf{x})$. Since the *a priori* probability of π_i is p_i, the *posterior* probability of π_i by Bayes theorem is

$$P(\pi_i \mid \mathbf{x}) = \frac{P(\pi_i, \mathbf{x})}{P(\mathbf{x})} = \frac{p_i f_i(\mathbf{x})}{p_1 f_1(\mathbf{x}) + p_2 f_2(\mathbf{x})} \quad i = 1,2$$

If we assign an observation to π_1 when $P(\pi_1 \mid \mathbf{x}) > P(\pi_2 \mid \mathbf{x})$, this is equivalent to the rule that minimizes the total probability of misclassification.

3. Some special situations occur when the populations have multivariate normal distributions ; each with parameters mean vector μ_i and covariance matrix Σ_i. Fisher's *linear discriminant function* can be used in the case of $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$ since it was developed under the assumption that the populations have a common covariance matrix. It corresponds to a particular case of the minimum expected cost of misclassification rule.

   The classification rules are more complicated when the population covariance matrices are unequal. This situation leads to the formulation of the *quadratic classification rules*.

(d) After a discriminant function has been calculated, its performance should be evaluated. Three questions are of major importance :

   1. Are the observed between-group differences statistically significant ? This determines if there is any hope of classifying future observations using the given variables. If not, then we might as well try to find better variables.

   2. If the differences are significant, are all of the variables needed in the discriminant functions ? Is a subset of the variables sufficient for discrimination ?

   3. How will the discriminant function perform on future samples ? This involves estimating the error rates of the given discriminant function.

3.3 Fisher's Method For Two Populations

3.3.1 Fisher's linear discriminant function

Let π_1 and π_2 be two classes of objects. The objects will be separated or classified on the basis of measurements on p associated random variables $\mathbf{X}' = [X_1, X_2, \cdots, X_p]$. The observed values of $\mathbf{X}'$, denoted by $\mathbf{x}'$, differ to some extent from one class to the other. The totality of values from the first class is the population of $\mathbf{x}$ values for π_1 and those from the second class is the population of $\mathbf{x}$ values for π_2. These two populations can be described by probability density functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$, and, consequently, we can talk of assigning observations to populations or objects to classes interchangeably.

Fisher's idea was to transform the multivariate observations $\mathbf{x}$ to univariate observations y such that the y's derived from populations π_1 and π_2 were separated as much as possible. Fisher suggested taking linear combinations of $\mathbf{x}$ to create the y's because they are simple functions of $\mathbf{x}$ and are easily handled mathematically.

The analysis begins by defining

$\mu_1 = E(\mathbf{X} \mid \pi_1)$ = vector of expected values of a multivariate observation from π_1

$\mu_2 = E(\mathbf{X} \mid \pi_2)$ = vector of expected values of a multivariate observation from π_2

$$(3.1)$$

and assuming the covariance matrix

$$\Sigma = E(\mathbf{X} - \mu_i)(\mathbf{X} - \mu_i)', \quad i = 1, 2 \tag{3.2}$$

is the same for both populations. The assumption of a common covariance matrix is somewhat critical. It is often violated in practice. One then considers the linear combination

$$Y_{(1\times 1)} = \ell'_{(1\times p)}\mathbf{X}_{(p\times 1)} \tag{3.3}$$

It can be shown that the random variable Y has a mean

$$\mu_{1Y} = E(Y \mid \pi_1) = E(\ell'\mathbf{X} \mid \pi_1) = \ell'\mu_1$$

or

$$\mu_{2Y} = E(Y \mid \pi_2) = E(\ell'\mathbf{X} \mid \pi_2) = \ell'\mu_2 \tag{3.4}$$

depending on the underlying population, but its variance

$$\sigma_Y^2 = \text{Var}(\ell'\mathbf{X}) = \ell'\text{Cov}(\mathbf{X})\ell = \ell'\mathbf{\Sigma}\ell \tag{3.5}$$

is the same for both populations.

The best linear combination is derived from the ratio

$$\begin{aligned}\frac{\begin{pmatrix}\text{Squared distance} \\ \text{between means of } Y\end{pmatrix}}{(\text{Variance of } Y)} &= \frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} = \frac{(\ell'\mu_1 - \ell'\mu_2)^2}{\ell'\mathbf{\Sigma}\ell} \\ &= \frac{\ell'(\mu_1 - \mu_2)(\mu_1 - \mu_2)'\ell}{\ell'\mathbf{\Sigma}\ell} = \frac{(\ell'\delta)^2}{\ell'\mathbf{\Sigma}\ell}\end{aligned} \tag{3.6}$$

where $\delta = (\mu_1 - \mu_2)$ is the difference in mean vectors. The $p \times p$ matrix $\delta\delta' = (\mu_1 - \mu_2)(\mu_1 - \mu_2)'$ contains the squares and cross-products of the component differences between the means of populations π_1 and π_2. Fisher's linear combination coefficients $\ell' = [\ell_1, \ell_2, \cdots, \ell_p]$ are those that maximize the ratio in (3.6). This ratio is maximized by the choice

$$\ell = c\mathbf{\Sigma}^{-1}\delta = c\mathbf{\Sigma}^{-1}(\mu_1 - \mu_2)$$

for any $c \neq 0$. Choosing $c = 1$ produces the linear combination

$$Y = \ell'\mathbf{X} = (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{X} \tag{3.7}$$

which is known as *Fisher's linear discriminant function* [Fisher, 1936].

The maximum of the ratio is given by

$$\max_{\ell} \frac{(\ell'\delta)^2}{\ell'\mathbf{\Sigma}\ell} = \delta'\mathbf{\Sigma}^{-1}\delta \tag{3.8}$$

The linear combination (3.7) can be used as a classification device by defining $y_0 = (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x}_0$ as the value of the discriminant function for a new observation $\mathbf{x}_0$ and letting

$$\begin{aligned} m &= \frac{1}{2}(\mu_{1Y} + \mu_{2Y}) \\ &= \frac{1}{2}(\ell'\mu_1 + \ell'\mu_2) \\ &= \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \end{aligned} \tag{3.9}$$

be the midpoint between the two univariate population means. It can be shown that

$$E(Y_0 \mid \pi_1) - m \geq 0$$

and

$$E(Y_0 \mid \pi_2) - m < 0 \tag{3.10}$$

That is, if $\mathbf{X_0}$ is from π_1, Y_0 is expected to be larger than the midpoint. If $\mathbf{X_0}$ is from π_2, Y_0 is expected to be smaller than the midpoint. Thus the classification rule is :

$$\begin{aligned} \text{Allocate } \mathbf{x}_0 \text{ to } \pi_1 \text{ if } y_0 &= (\mu_1 - \mu_2)'\Sigma^{-1}\mathbf{x}_0 \geq m \\ \text{Allocate } \mathbf{x}_0 \text{ to } \pi_2 \text{ if } y_0 &= (\mu_1 - \mu_2)'\Sigma^{-1}\mathbf{x}_0 < m \end{aligned} \tag{3.11}$$

Alternatively, we can substract m from y_0 and compare the result with zero. In this case, the rule becomes :

$$\begin{aligned} \text{Allocate } \mathbf{x}_0 \text{ to } \pi_1 \text{ if } y_0 - m &\geq 0 \\ \text{Allocate } \mathbf{x}_0 \text{ to } \pi_2 \text{ if } y_0 - m &< 0 \end{aligned} \tag{3.12}$$

[Johnson *et al*, 1988].

3.3.2 Fisher's sample linear discriminant function

The population quantities μ_1, μ_2, and Σ are rarely known. Therefore, the rules in (3.11) and (3.12) cannot be implemented unless ℓ and m can be estimated from observations that have already been correctly classified.

Suppose, then, that we have n_1 observations of the multivariate random variable $\mathbf{X}' = [X_1, X_2, \cdots, X_p]$ from π_1 and n_2 measurements of this quantity from π_2. The respective data matrices are

$$\begin{aligned} \mathbf{X}_1^{(p \times n_1)} &= [\mathbf{x}_{11}, \mathbf{x}_{12}, \cdots, \mathbf{x}_{1n_1}] \\ \mathbf{X}_2^{(p \times n_2)} &= [\mathbf{x}_{21}, \mathbf{x}_{22}, \cdots, \mathbf{x}_{2n_2}] \end{aligned} \tag{3.13}$$

where $\mathbf{X}_1$ is a $(p \times n_1)$ matrix and $\mathbf{X}_2$ is a $(p \times n_2)$ matrix.

$$R_2 : \exp[-\frac{1}{2}(\mathbf{x} - \mu_1)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_2)] < \left[\frac{c(1 \mid 2)}{c(2 \mid 1)}\right]\left[\frac{p_2}{p_1}\right] \tag{3.31}$$

Since the quantities in (3.31) are nonnegative for all $\mathbf{x}$, we can take their natural logarithms and preserve the order of the inequalities. Moreover,

$$\begin{aligned} & -\frac{1}{2}(\mathbf{x} - \mu_1)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_2) \\ & = (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \end{aligned} \tag{3.32}$$

and, consequently,

$$\begin{aligned} R_1 &: (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \geq \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \\ R_2 &: (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) < \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \end{aligned} \tag{3.33}$$

For the populations π_1 and π_2 described by multivariate normal densities of the form (3.30), the allocation rule that minimizes the ECM is given by :

Allocate $\mathbf{x}_0$ to π_1 if

$$(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x}_0 - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \geq \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{3.34}$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Comparing the minimum ECM rule in (3.34) with the population analog of Fisher's method summarized by (3.9), (3.11), and (3.12), it is clear the two procedures are identical when

$$\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right) = 1$$

since $\ln(1) = 0$.

Unknown population parameters

The population quantities $\mu_1, \mu_2,$ and $\mathbf{\Sigma}$ are unknown in most practical situations, so the rule (3.34) must be modified by replacing the population parameters by their sample counterparts. Substituting $\mathbf{x}_1$ for μ_1, $\mathbf{x}_2$ for μ_2, and $\mathbf{S}_{pooled}$ for $\mathbf{\Sigma}$ in (3.34) gives the following "sample" classification rule :

The sample linear discriminant function in (3.16) has the following "optimal" property, that is, the linear combination $y = \hat{\ell}'\mathbf{x} = (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}_{pooled}^{-1}\mathbf{x}$ maximizes the ratio

$$\frac{\begin{pmatrix}\text{Squared distance}\\ \text{between sample means of } Y\end{pmatrix}}{(\text{Sample variance of } Y)} = \frac{(y_1 - y_2)^2}{s_y^2} = \frac{(\hat{\ell}'\mathbf{x}_1 - \hat{\ell}'\mathbf{x}_2)^2}{\hat{\ell}'\mathbf{S}_{pooled}\hat{\ell}} = \frac{(\hat{\ell}'\mathbf{d})^2}{\hat{\ell}'\mathbf{S}_{pooled}\hat{\ell}} \tag{3.19}$$

where $\mathbf{d} = (\mathbf{x}_1 - \mathbf{x}_2)$.

The ratio in (3.19) is the sample analog of the ratio in (3.6). Also, s_y^2 may be calculated as

$$s_y^2 = \left(\frac{\sum_{j=1}^{n_1}(y_{1j} - y_1)^2 + \sum_{j=1}^{n_2}(y_{2j} - y_2)^2}{n_1 + n_2 - 2}\right) \tag{3.20}$$

with $y_{1j} = \hat{\ell}'\mathbf{x}_{1j}$ and $y_{2j} = \hat{\ell}'\mathbf{x}_{2j}$ [Johnson *et al*, 1988].

3.3.3 Hypothesis testing

The maximum value of the population ratio in (3.6) is, from (3.8), $\delta'\Sigma^{-1}\delta = (\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$. This is the squared distance, Δ^2, between two populations. The maximum of the sample ratio in (3.19) is given by setting $\hat{\ell} = \mathbf{S}_{pooled}^{-1}(\mathbf{x}_1 - \mathbf{x}_2)$. Thus

$$\max_{\hat{\ell}} \frac{(\hat{\ell}'\mathbf{d})^2}{\hat{\ell}'\mathbf{S}_{pooled}\hat{\ell}} = \mathbf{d}'\mathbf{S}_{pooled}^{-1}\mathbf{d} = (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}_{pooled}^{-1}(\mathbf{x}_1 - \mathbf{x}_2) = D^2 \tag{3.21}$$

where D^2 is the sample squared distance known as Mahalanobis' distance [Lachenbruch, 1975].

For two populations, D^2 can be used, in certain situations, to test whether the population means μ_1 and μ_2 differ significantly. A test for differences in mean vectors can also be viewed as a test for the "significance" of the separation that can be achieved.

Suppose the population π_1 and π_2 are multivariate normal *with a common covariance matrix* Σ. Then, a test of $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 \neq \mu_2$ is accomplished by referring

$$\left(\frac{n_1 + n_2 - p - 1}{(n_1 + n_2 - 2)p}\right)\left(\frac{n_1 n_2}{n_1 + n_2}\right)D^2$$

to an F-distribution with $\nu_1 = p$ and $\nu_2 = n_1 + n_2 - p - 1$ d.f. If H_0 is rejected, we can conclude the separation between the two populations π_1 and π_2 is significant [Johnson *et al*, 1988].

3.4 The General Problem of Classification

3.4.1 Bayes Procedure for Two Populations

Suppose an individual is an observation from either population π_1 or population π_2. The classification of an observation depends on the vector of measurements $\mathbf{x}' = (x_1, x_2, \cdots, x_p)$ on that individual. We set up a rule that if an individual is characterized by certain sets of values of $x_1, x_2, \cdots, x_p$ it will be classified as from π_1 ; if it has other values it is classified as from π_2.

We can think of an observation as a point in a p-dimensional space. We divide this space into two regions. If the observation falls in R_1, we classify it as coming from population π_1, and if it falls in R_2 we classify it as coming from population π_2 [Anderson, 1958].

Classification rules cannot usually provide an error-free method of assignment. This is because there may not be a clear distinction between the measured characteristics of the populations ; that is, the groups may overlap. It is then possible, for example, to incorrectly classify a π_2 objects as belonging to π_1 or a π_1 object as belonging to π_2.

A good classification procedure should result in few misclassifications. In other words, the chances, or probabilities, of misclassification should be small. There are additional features that an "optimal" classification rule should possess :

(a) It may be that one class or population has a greater likelihood of occurence than another because one of the two populations is relatively much larger than the other. An optimal classification rule should take these *prior probabilities of occurence* into account.

(b) Another aspect of classification is cost. Suppose that classifying a π_1 object as belonging to π_2 represents a more serious error than classifying a π_2 object as belonging to π_1. Then one should be cautious about making the former assignment. An optimal classification procedure should, whenever possible, account for the costs associated with misclassification

[Johnson *et al*, 1988] [Anderson, 1958].

Let $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ be the probability density functions associated with the $p \times 1$ vector random variable $\mathbf{X}$ for the populations π_1 and π_2, respectively. An object, with associated measurements $\mathbf{x}$, *must* be assigned to either π_1 or π_2. Let Ω be the sample space ; that is, the collection of all possible observations $\mathbf{x}$. Let R_1 be that set of $\mathbf{x}$ values for which we classify objects as π_1 and $R_2 = \Omega - R_1$ be the remaining $\mathbf{x}$ values for which we classify objects as π_2. Since every objects must be assigned to one and only one of the populations, the sets R_1 and R_2 are mutually exclusive and exhaustive.

The conditional probability, $P(2 \mid 1)$, of classifying an object as π_2 when, in fact, it is from π_1 is

$$P(2 \mid 1) = P(\mathbf{X} \in R_2 \mid \pi_1) = \int_{R_2 = \Omega - R_1} f_1(\mathbf{x}) d\mathbf{x} \tag{3.22}$$

Similarly, the conditional probability, $P(1 \mid 2)$, of classifying an object as π_1 when it is really from π_2 is

$$P(1 \mid 2) = P(\mathbf{X} \in R_1 \mid \pi_2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} \tag{3.23}$$

The integral sign in (3.22) represents the volume formed by the density function $f_1(\mathbf{x})$ over the region R_2. Similarly, the integral sign in (3.23) represents the volume formed by $f_2(\mathbf{x})$ over the region R_1.

Let p_1 be the *prior* probability of π_1 and p_2 be the *prior* probability of π_2, where $p_1 + p_2 = 1$. The overall probabilities of correctly or incorrectly classifying objects can be derived as the product of the prior and conditional classification probabilities :

$$\begin{aligned}
P(\text{correctly classified as } \pi_1) &= P(\text{observation comes from } \pi_1 \text{ and} \\
&\qquad \text{is correctly classified as } \pi_1) \\
&= P(\mathbf{X} \in R_1 \mid \pi_1) P(\pi_1) = P(1 \mid 1) p_1 \\
P(\text{misclassified as } \pi_1) &= P(\text{observation comes from } \pi_2 \text{ and} \\
&\qquad \text{is misclassified as } \pi_1)
\end{aligned}$$

$$
\begin{aligned}
&= P(\mathbf{X} \in R_1 \mid \pi_2)P(\pi_2) = P(1 \mid 2)p_2 \\
P(\text{correctly classified as } \pi_2) &= P(\text{observation comes from } \pi_2 \text{ and} \\
&\qquad \text{is correctly classified as } \pi_2) \\
&= P(\mathbf{X} \in R_2 \mid \pi_2)P(\pi_2) = P(2 \mid 2)p_2 \\
P(\text{misclassified as } \pi_2) &= P(\text{observation comes from } \pi_1 \text{ and} \\
&\qquad \text{is misclassified as } \pi_2) \\
&= P(\mathbf{X} \in R_2 \mid \pi_1)P(\pi_1) = P(2 \mid 1)p_1
\end{aligned}
\tag{3.24}
$$

Classification schemes are often evaluated in terms of their misclassification probabilities, but this ignores misclassification cost. A rule that ignores costs may cause problems.

The costs of misclassification can be defined by a cost matrix.

		Classify as	
		π_1	π_2
True	π_1	0	$c(2 \mid 1)$
population	π_2	$c(1 \mid 2)$	0

(3.25)

The costs are :

(1) zero for correct classification,

(2) $c(1 \mid 2)$ when an observation from π_2 is incorrectly classified as π_1, and

(3) $c(2 \mid 1)$ when a π_1 observation is incorrectly classified as π_2.

For any rule, the average, or *expected cost of misclassification* (ECM) is provided by multiplying the off-diagonal entries in (3.25) by their probabilities of occurence, obtained from (3.24). Consequently,

$$\text{ECM} = c(2 \mid 1)P(2 \mid 1)p_1 + c(1 \mid 2)P(1 \mid 2)p_2 \tag{3.26}$$

A reasonable classification rule should have an ECM as small, or nearly as small, as possible. The assignment regions R_1 and R_2 must be chosen so that the ECM is as small as possible [Johnson *et al*, 1988].

A procedure that minimizes (3.26) for a given p_1 and p_2 is called a Bayes procedure [Anderson, 1958]. The regions R_1 and R_2 that minimize the ECM are defined by the values **x** for which the following inequalities hold.

$$R_1 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \left[\frac{c(1 \mid 2)}{c(2 \mid 1)}\right] \left[\frac{p_2}{p_1}\right]$$

$$\left[\begin{array}{c} \text{Density} \\ \text{ratio} \end{array} \right] \geq \left[\begin{array}{c} \text{cost} \\ \text{ratio} \end{array} \right] \left[\begin{array}{c} \text{prior} \\ \text{probability} \\ \text{ratio} \end{array} \right] \qquad (3.27)$$

$$R_2 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \left[\frac{c(1 \mid 2)}{c(2 \mid 1)}\right] \left[\frac{p_2}{p_1}\right]$$

$$\left[\begin{array}{c} \text{Density} \\ \text{ratio} \end{array} \right] < \left[\begin{array}{c} \text{cost} \\ \text{ratio} \end{array} \right] \left[\begin{array}{c} \text{prior} \\ \text{probability} \\ \text{ratio} \end{array} \right]$$

The appearance of ratios in the definition of the optimal classification regions is significant. Often it is much easier to specify the ratios than their component parts [Anderson, 1958].

Some special cases

It is also interesting to consider the classification regions defined in (3.27) for some special cases :

(a) $(p_2/p_1) = 1$ (equal prior probabilities)

$$R_1 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{c(1 \mid 2)}{c(2 \mid 1)} \; : \; R_2 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{c(1 \mid 2)}{c(2 \mid 1)}$$

(b) $[c(1 \mid 2)/c(2 \mid 1)] = 1$ (equal misclassification cost)

$$R_1 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq \frac{p_2}{p_1} \; ; \; R_2 \; : \; \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < \frac{p_2}{p_1} \qquad (3.28)$$

This case is the same with minimizing the *total probability of misclassification* (TPM) or allocating a new observation $\mathbf{x}_0$ to the population with the largest "posterior" probability $P(\pi_i \mid \mathbf{x}_0)$, where

$$\begin{aligned}
P(\pi_1 \mid \mathbf{x}_0) &= \frac{P(\pi_1 \text{occurs and observe } \mathbf{x}_0)}{P(\text{observe } \mathbf{x}_0)} \\
&= \frac{P(\text{observe } \mathbf{x}_0 \mid \pi_1)P(\pi_1)}{P(\text{observe } \mathbf{x}_0 \mid \pi_1)P(\pi_1) + P(\text{observe } \mathbf{x}_0 \mid \pi_2)P(\pi_2)} \\
&= \frac{p_1 f_1(\mathbf{x}_0)}{p_1 f_1(\mathbf{x}_0) + p_2 f_2(\mathbf{x}_0)} \\
P(\pi_2 \mid \mathbf{x}_0) &= 1 - P(\pi_1 \mid \mathbf{x}_0) \\
&= \frac{p_2 f_2(\mathbf{x}_0)}{p_1 f_1(\mathbf{x}_0) + p_2 f_2(\mathbf{x}_0)} \qquad (3.29)
\end{aligned}$$

(c) $[p_2/p_1] = [c(1 \mid 2)/c(2 \mid 1)] = 1$ or $[p_2/p_1] = 1/[c(1 \mid 2)/c(2 \mid 1)]$ (equal prior probabilities and equal misclassification costs)

$$R_1 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} \geq 1 \ \ ; \ \ R_2 \ : \ \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} < 1$$

[Johnson *et al*, 1988].

3.5 Classification with Two Multivariate Normal Populations

We now assume $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are multivariate normal densities ; the first with mean vector μ_1 and covariance matrix $\mathbf{\Sigma}_1$ and the second with mean vector μ_2 and covariance matrix $\mathbf{\Sigma}_2$.

Let the joint densities of $\mathbf{X}' = [X_1, X_2, \cdots, X_p]$ for populations π_1 and π_2 be given by

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \mid \mathbf{\Sigma} \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_i)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_i)\right] \quad \text{for} \ \ i = 1, 2 \qquad (3.30)$$

3.5.1 The case of $\mathbf{\Sigma_1} = \mathbf{\Sigma_2} = \mathbf{\Sigma}$

Known population parameters

Suppose the population parameters μ_1, μ_2, and $\mathbf{\Sigma}$ are known.

After cancellation of the term $(2\pi)^{p/2} \mid \mathbf{\Sigma} \mid^{1/2}$ and a rearrangement of the exponents in the multivariate normal densities in (3.30), the minimum ECM regions in (3.27) become

$$R_1 : \exp[-\frac{1}{2}(\mathbf{x} - \mu_1)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_2)] \ \geq \ \left[\frac{c(1 \mid 2)}{c(2 \mid 1)}\right]\left[\frac{p_2}{p_1}\right]$$

$$R_2 : \exp[-\frac{1}{2}(\mathbf{x} - \mu_1)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_2)] \quad < \quad \left[\frac{c(1 \mid 2)}{c(2 \mid 1)}\right]\left[\frac{p_2}{p_1}\right] \tag{3.31}$$

Since the quantities in (3.31) are nonnegative for all $\mathbf{x}$, we can take their natural logarithms and preserve the order of the inequalities. Moreover,

$$\begin{aligned} - & \frac{1}{2}(\mathbf{x} - \mu_1)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_1) + \frac{1}{2}(\mathbf{x} - \mu_2)'\mathbf{\Sigma}^{-1}(\mathbf{x} - \mu_2) \\ = & \ (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \end{aligned} \tag{3.32}$$

and, consequently,

$$\begin{aligned} R_1 : (\mu_1 - \mu_?)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \quad \geq \quad \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \\ R_2 : (\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \quad < \quad \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \end{aligned} \tag{3.33}$$

For the populations π_1 and π_2 described by multivariate normal densities of the form (3.30), the allocation rule that minimizes the ECM is given by :

Allocate $\mathbf{x}_0$ to π_1 if

$$(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}\mathbf{x}_0 - \frac{1}{2}(\mu_1 - \mu_2)'\mathbf{\Sigma}^{-1}(\mu_1 + \mu_2) \geq \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{3.34}$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Comparing the minimum ECM rule in (3.34) with the population analog of Fisher's method summarized by (3.9), (3.11), and (3.12), it is clear the two procedures are identical when

$$\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right) = 1$$

since $\ln(1) = 0$.

Unknown population parameters

The population quantities μ_1, μ_2, and $\mathbf{\Sigma}$ are unknown in most practical situations, so the rule (3.34) must be modified by replacing the population parameters by their sample counterparts. Substituting $\mathbf{x}_1$ for μ_1, $\mathbf{x}_2$ for μ_2, and $\mathbf{S}_{pooled}$ for $\mathbf{\Sigma}$ in (3.34) gives the following "sample" classification rule :

The Estimated Minimum ECM Rule for Two Normal Population :

Allocate $\mathbf{x}_0$ to π_1 if

$$(\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}_{pooled}^{-1}\mathbf{x}_0 - \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}_{pooled}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \geq \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{3.35}$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

The first term, $y = (\mathbf{x}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{pooled}^{-1}\mathbf{x}$ in (3.35) is the linear function obtained by Fisher that maximizes the univariate "between" samples variability relative to the "within" samples variability [see (3.19)]. The entire expression

$$\begin{aligned} w &= (\mathbf{x}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{pooled}^{-1}\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_{pooled}^{-1}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2) \\ &= (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}_{pooled}^{-1}[\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_2)] \end{aligned} \tag{3.36}$$

is often called *Anderson's classification function (statistic).* [Anderson, 1958]

Once again, if $[(c(1\mid 2)/c(2\mid 1))(p_2/p_1)] = 1$ so that $\ln[(c(1\mid 2)/c(2\mid 1))(p_2/p_1)] = 0$, Rule (3.35) is comparable to Rule (3.18) based on Fisher's linear discriminant function. Thus, provided the two normal populations have the same covariance matrix, Fisher's classification rule is equivalent to the minimum ECM rule with equal prior probabilities and equal costs of misclassification.

Once parameter estimates are inserted for the corresponding unknown population quantities, there is no assurance the resulting rule will minimize the expected cost of misclassification in a particular application. This is because the optimal rule in (3.34) was derived, assuming the multivariate normal densities $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ were known completely. Expression (3.35) is simply an estimate of the optimal rule. However, it seems reasonable to expect that it should perform well **if the sample sizes are large**. As the sample sizes increase, $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{S}_{pooled}$ become, with probability approaching 1, indistinguishable from μ_1, μ_2, and Σ, respectively.

To summarize, if the data appear to be multivariate normal, [1] the classification statistic w in (3.36) can be calculated for each new observation $\mathbf{x}_0$. These observations are classified

[1] At the very least the marginal frequency distributions of the observations on each variable can be checked for normality. This must be done for the samples from both populations. Often some variables must be transformed in order to make them more "normal looking". [Johnson *et al*, 1988]

by comparing the values of w with the value of $\ln[(c(1 \mid 2)/c(2 \mid 1))(p_2/p_1)]$ as in (3.35). [Johnson *et al*, 1988]

3.5.2 The case of $\Sigma_1 \neq \Sigma_2$

The classification rules are more complicated when the population covariance matrices are unequal.

Known population parameters

Substituting multivariate normal densities with different covariance matrices into (3.27) gives, after taking natural logarithms and simplifying, the classification regions

$$\begin{aligned} R_1 : -\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})\mathbf{x} - k &\geq \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \\ R_2 : -\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x} + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})\mathbf{x} - k &< \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \end{aligned} \tag{3.37}$$

where

$$k = \frac{1}{2}\ln\left(\frac{|\Sigma_1|}{|\Sigma_2|}\right) + \frac{1}{2}(\mu_1'\Sigma_1^{-1}\mu_1 - \mu_2'\Sigma_2^{-1}\mu_2) \tag{3.38}$$

The classification regions are defined by *quadratic* functions of $\mathbf{x}$. When $\Sigma_1 = \Sigma_2$, the quadratic term, $-\frac{1}{2}\mathbf{x}'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x}$, disappears and the regions defined by (3.37) reduce to those defined by (3.33).

The allocation rule that minimizes the expected cost of misclassification is given by :

Allocate $\mathbf{x}_0$ to π_1 if

$$-\frac{1}{2}\mathbf{x}_0'(\Sigma_1^{-1} - \Sigma_2^{-1})\mathbf{x}_0 + (\mu_1'\Sigma_1^{-1} - \mu_2'\Sigma_2^{-1})\mathbf{x}_0 - k \geq \ln\left[\left(\frac{c(1 \mid 2)}{c(2 \mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Unknown population parameters

In practice, the classification rule above is implemented by substituting the sample quantities $\mathbf{x}_1, \mathbf{x}_2, \mathbf{S}_1$, and $\mathbf{S}_2$ (see (3.14)) for μ_1, μ_2, Σ_1, and Σ_2, respectively. [2] This gives the following "sample" classification rule :

The Quadratic Classification Rule

(Two normal populations with unequal covariance matrices)

Allocate $\mathbf{x}_0$ to π_1 if

$$-\frac{1}{2}\mathbf{x}_0'(\mathbf{S}_1^{-1} - \mathbf{S}_2^{-1})\mathbf{x}_0 + (\mathbf{x}_1'\mathbf{S}_1^{-1} - \mathbf{x}_2'\mathbf{S}_2^{-1})\mathbf{x}_0 - k \geq \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right] \tag{3.39}$$

Allocate $\mathbf{x}_0$ to π_2 otherwise.

Classification with quadratic functions is awkward in more than two dimensions and can lead to strange results. This is particularly true when the data are not (essentially) multivariate normal.

If the data are not multivariate normal, two options are available. First, the nonnormal data can be transformed to data more nearly normal and a test for the equality of covariance matrices can be conducted to see if the linear rule (3.35) or the quadratic rule (3.39) is appropriate. The tests for covariance homogeneity are greatly affected by nonnormality. The conversion of nonnormal data to normal data must be done before this testing is carried out.

Second, we can use a linear (or quadratic) rule without worrying about the form of the parent populations and hope that it will work reasonably well. Fisher's procedure, for example, did not depend on the form of the parent populations, apart from the requirement of identical covariance structures. The moral is to always check the performance of any classification procedure. At the very least, this should be done with the data sets used to build the classifier. Ideally, there will be enough data available to provide "training" samples and "validation" samples. The training samples can be used to develop the classification

[2] The inequalities $n_1 > p$ and $n_2 > p$ must both hold for $\mathbf{S}_1^{-1}$ and $\mathbf{S}_2^{-1}$ to exist. These quantities are used in place of Σ_1^{-1} and Σ_2^{-1}, respectively, in the sample analog (3.39).

function and the validation samples can be used to evaluate its performance [Johnson *et al.*, 1988].

3.6 Evaluating Classification Functions

3.6.1 Known population parameters

One important way of judging the performance of any classification procedure is to calculate its "error rates," or misclassification probabilities. When the forms of the parent populations are known completely, misclassification probabilities can be calculated with relative ease. Because parent populations are rarely known, we shall concentrate on the error rates associated with the sample classification function. Once this classification function is constructed, a measure of its performance in *future* samples is of interest.

The ***total probability of misclassification*** (TPM) is

$$\mathrm{TPM} = p_1 \int_{R_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x})d\mathbf{x}$$

The smallest value of this quantity, obtained by a judicious choice of R_1 and R_2, is called the optimum error rate (OER).

$$\text{Optimum error rate (OER)} = p_1 \int_{R_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{R_1} f_2(\mathbf{x})d\mathbf{x} \tag{3.40}$$

where R_1 and R_2 are determined by case (b) in (3.28).

Thus the OER is the error rate for the minimum TPM classification rule [Johnson *et al.*, 1988].

3.6.2 Unknown population parameters

The optimum error rate can be calculated when the population density functions are known. If, as is usually the case, certain population parameters appearing in allocation rules must be estimated from the sample, then the evaluation of error rates is not straightforward.

The performance of *sample* classification functions can, in principle, be evaluated by calculating the actual error rate (AER),

$$\text{Actual error rate (AER)} = p_1 \int_{\hat{R}_2} f_1(\mathbf{x})d\mathbf{x} + p_2 \int_{\hat{R}_1} f_2(\mathbf{x})d\mathbf{x} \tag{3.41}$$

where $\hat{R}_1$ and $\hat{R}_2$ represent the classification regions determined by samples of size n_1 and n_2, respectively. For example, if the classification function in (3.36) is employed, the regions $\hat{R}_1$ and $\hat{R}_2$ are defined by the set of x's for which the following inequalities are satisfied.

$$\hat{R}_1 : (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}^{-1}_{pooled}\mathbf{x} - \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}^{-1}_{pooled}(\mathbf{x}_1 + \mathbf{x}_2) \;\geq\; \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$
$$\hat{R}_2 : (\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}^{-1}_{pooled}\mathbf{x} - \frac{1}{2}(\mathbf{x}_1 - \mathbf{x}_2)'\mathbf{S}^{-1}_{pooled}(\mathbf{x}_1 + \mathbf{x}_2) \;<\; \ln\left[\left(\frac{c(1\mid 2)}{c(2\mid 1)}\right)\left(\frac{p_2}{p_1}\right)\right]$$

The AER indicates how the sample classification function will perform in future samples. Like the optimal error rate, it cannot, in general, be calculated because it depends on the unknown density functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$. However, an estimate of a quantity related to the actual error rate can be calculated [Johnson *et al*, 1988].

3.6.3 General case

There is a measure of performance that does not depend on the form of the parent populations and that can be calculated for *any* classification procedure. This measure, called the *apparent error rate* (APER), is defined as the fraction of observations in the *training* sample that are misclassified by the sample classification function.

The apparent error rate can be easily calculated from the *confusion matrix*, which shows actual versus predicted group membership. The confusion matrix is built by applying the values of the variables in the measurement vectors of known groups, and by recording the number of correct and incorrect classifications into a two-dimensional table of classification. For n_1 observations from π_1 and n_2 observations from π_2, the confusion matrix has the form

		Predicted membership		
		π_1	π_2	
Actual	π_1	n_{1C}	$n_{1M} = n_1 - n_{1C}$	n_1
Membership	π_2	$n_{2M} = n_2 - n_{2C}$	n_{2C}	n_2

(3.42)

where

n_{1C} = number of π_1 items correctly classified as π_1 items

n_{1M} = number of π_1 items misclassified as π_2 items

n_{2C} = number of π_2 items correctly classified as π_2 items

n_{2M} = number of π_2 items misclassified as π_1 items

The apparent error rate is then

$$\text{APER} = \frac{n_{1M} + n_{2M}}{n_1 + n_2} \tag{3.13}$$

which is recognized as the *proportion* of items in the training that are misclassified.

The APER is intuitively appealing and easy to calculate. Unfortunately, it tends to underestimate the AER, and this problem does not disappear unless the sample sizes n_1 and n_2 are very large. This optimistic estimate occurs because the data used to build the classification function are also used to evaluate it [Johnson *et al*, 1988].

3.6.4 Partition of Total Sample and Jackknifing

Error-rate estimates can be constructed that are better than the apparent error rate, remain relatively easy to calculate, and do not require distributional assumptions. One procedure is to split the total sample into a training sample and a validation sample. The training sample is used to construct the classification function and the validation sample is used to evaluate it. The error rate is determined by the proportion misclassified in the validation sample. Although this method overcomes the bias problem by not using the same data to both build and judge the classification function, it suffers from two main defects.

1. It requires large samples.

2. The function evaluated is not the function of interest. Ultimately, almost *all* of the data must be used to construct the classification function. If not, valuable information may be lost

[Johnson *et al*, 1988].

A second approach that seems to work well is called Lachenbruch's "holdout" procedure [Lachenbruch *et al*, 1968], and is sometimes referred to as *jackknifing* :

1. Start with the π_1 group of observations. Omit one observation from this group and develop a classification function based on the remaining $n_1 - 1$, n_2 observations.

2. Classify the "holdout" observation using the function constructed in Step 1.

3. Repeats Steps 1 and 2 until all of the π_1 observations are classified. Let $n_{1M}^{(H)}$ be the number of holdout (H) observations misclassified in this group.

4. Repeat Steps 1 through 3 for the π_2 observations. Let $n_{2M}^{(H)}$ be the number of holdout observations misclassified in this group.

Estimates $\hat{P}(2 \mid 1)$ and $\hat{P}(1 \mid 2)$ of the conditional misclassification probabilities in (3.22) and (3.23) are then given by

$$\hat{P}(2 \mid 1) = \frac{n_{1M}^{(H)}}{n_1} \qquad (3.44)$$
$$\hat{P}(1 \mid 2) = \frac{n_{2M}^{(H)}}{n_2}$$

and the total proportion misclassified, $(n_{1M}^{(H)} + n_{2M}^{(H)})/(n_1 + n_2)$ is, for moderate samples, a nearly unbiased estimated of the *expected* actual error rate, $E(\text{AER})$.

$$\hat{E}(\text{AER}) = \frac{(n_{1M}^{(H)} + n_{2M}^{(H)})}{(n_1 + n_2)} \qquad (3.45)$$

Lachenbruch's holdout method is computationally feasible when used in conjunction with the linear classification statistics in (3.16) and (3.36). It is offered as an option in some readily available discriminant analysis computer programs [Dixon *et al*, 1988].

CHAPTER 4

RESEARCH METHODS AND MATERIALS

4.1 Introduction

As mentioned in Section 1.4, the first objective of this research is to utilize variation as the basis for identifying groups of characters or letters in crypto texts through discrimination and classification processes developed by using samples of source texts. The characters are extracted from texts or articles written in English. This research uses a collection of *four hundred and fifty three files* of articles published in *Computing Canada*. The purpose of using articles from only one journal is to make *within-group* variation as small as possible. This is based on the assumption that the articles will have the same style of writing. These files of articles are converted so that they will have only capital letters A to Z, and Blanks or Spaces. All special and punctuation characters are removed by a computer program written in a high-level programming language, such as Pascal or FORTRAN.

The second objective of this research is to increase the amount of information and variation by generating more variables - which have an *invariance* feature between source and crypto texts - from the sample of source texts by using the digraph structure or higher structure of the source texts. All variables are generated by a computer program written in Pascal.

The use of Multivariate Discriminant Analysis is the third objective of this research. The theoretical basis of the analysis involving two populations is discussed in the previous chapter. This chapter discusses the expansion of aspects when the number of populations is greater than two, since this research involves twenty seven groups or populations ($g = 27$).

The first part of this chapter is about the Bayes procedure and Fisher's method for several populations. The second part is about the implementation of these procedures and methods in this research.

4.2 Bayes Procedure for Several Populations

In theory, the generalization of classification procedures from 2 to g $\geq$ 2 groups is straight forward. However, not much is known about the properties of the corresponding *sample* classification functions and, in particular, their error rates have not been fully investigated.

The approach in this section is to develop the theoretically optimal rules and then indicate the modifications required for real-world applications.

4.2.1 The Minimum Expected Cost of Misclassification Method

Let $f_i(\mathbf{x})$ be the density associated with population π_i, $i = 1, 2, \cdots, g$. [For the most part, we shall take $f_i(\mathbf{x})$ to be a multivariate normal density, but this is unnecessary for the development of the general theory]. Let

$$
\begin{aligned}
p_i &= \text{the prior probability of population } \pi_i, \quad i = 1, 2, \cdots, g \\
c(k \mid i) &= \text{the cost of allocating an item to } \pi_k \text{ when, in fact, it belongs to} \\
&\quad\; \pi_i, \quad \text{for } k, i = 1, 2, \cdots, g \\
&\quad\; \text{For } k = i, \; c(i \mid i) = 0.
\end{aligned}
$$

Finally, let R_k be the set of $\mathbf{x}$'s classified as π_k and

$$P(k \mid i) = P(\text{classify item as } \pi_k \mid \pi_i) = \int_{R_k} f_i(\mathbf{x}) d\mathbf{x}$$

for $k, i = 1, 2, \cdots, g$ with $P(i \mid i) = 1 - \sum_{k=1}^{g} P(k \mid i), k \neq i$.

The conditional expected cost of misclassifying an $\mathbf{x}$ from π_1 into π_2, or $\pi_3, \cdots$, or π_g is

$$
\begin{aligned}
\mathrm{ECM}(1) &= P(2 \mid 1)c(2 \mid 1) + P(3 \mid 1)c(3 \mid 1) + \cdots + P(g \mid 1)c(g \mid 1) \\
&= \sum_{k=2}^{g} P(k \mid 1)c(k \mid 1)
\end{aligned}
$$

This conditional expected cost occurs with prior probability p_1, the probability of π_1.

In a similar manner, we can obtain the conditional expected costs of misclassification, ECM(2),$\cdots$, ECM(g). Multiplying each conditional ECM by its prior probability and summing gives the overall ECM.

$$\begin{aligned}
\mathrm{ECM} &= p_1 ECM(1) + p_2 ECM(2) + \cdots + p_g ECM(g) \\
&= p_1 \left(\sum_{k=2}^{g} P(k \mid 1) c(k \mid 1) \right) + p_2 \left(\sum_{k=1, k \neq 2}^{g} P(k \mid 2) c(k \mid 2) \right) \\
&\quad + \cdots + p_g \left(\sum_{k=1}^{g-1} P(k \mid g) c(k \mid g) \right) \\
&= \sum_{i=1}^{g} p_i \left(\sum_{k=1, k \neq i}^{g} P(k \mid i) c(k \mid i) \right) \qquad (4.1)
\end{aligned}$$

Determining an optimal classification procedure amounts to choosing the mutually exclusive and exhaustive classification regions $R_1, R_2, \cdots, R_g$ such that (4.1) is a minimum. As a result, the classification regions that minimize the ECM (4 1) are defined by allocating $\mathbf{x}$ to that population π_k, $k = 1, 2, \cdots g$ for which

$$\sum_{i=1, i \neq k}^{g} p_i f_i(\mathbf{x}) c(k \mid i) \qquad (4.2)$$

is smallest. If a tie occurs, $\mathbf{x}$ can be assigned to any of the tied populations.

Suppose all the misclassification costs are equal. (Without loss of generality we can set them equal to 1.) Using the argument leading to (4.2) , we would allocate $\mathbf{x}$ to that population $\pi_k, k = 1, 2, \cdots, g$, for which

$$\sum_{i=1, i \neq k}^{g} p_i f_i(\mathbf{x}) \qquad (4.3)$$

is smallest. Now (4.3) will be smallest when the omitted term, $p_k f_k(\mathbf{x})$, is ***largest***. Consequently, when the misclassification costs are the same, the minimum expected cost of misclassification rule has the following simple form.

Minimum ECM Classification Rule with Equal Misclassification Costs :

Allocate $\mathbf{x}$ to π_k if

$$p_k f_k(\mathbf{x}) > p_i f_i(\mathbf{x}) \text{ for all } i \neq k \tag{4.4}$$

or, equivalently,

Allocate $\mathbf{x}$ to π_k if

$$\ln p_k f_k(\mathbf{x}) > \ln p_i f_i(\mathbf{x}) \text{ for all } i \neq k \tag{4.5}$$

It is interesting to note that the classification rule in (4.4) is identical to the one that maximizes the "posterior" probability, $P(\pi_k \mid \mathbf{x}) = P(\mathbf{x}$ comes from π_k given that $\mathbf{x}$ was observed), where

$$P(\pi_k \mid \mathbf{x}) = \frac{p_k f_k(\mathbf{x})}{\sum_{i=1}^{g} p_i f_i(\mathbf{x})} = \frac{(\text{prior}) \times (\text{likelihood})}{\Sigma[(\text{prior}) \times (\text{likelihood})]} \text{ for } k = 1, 2, \cdots, g \tag{4.6}$$

Equation (4.6) is the generalization of Equation (3.29) to $g \geq 2$ groups.

In general, the minimum ECM rules have three components : prior probabilities, misclassification costs, and density functions. These components must be specified (or estimated) before the rules can be implemented [Anderson, 1958].

4.2.2 Classification with Normal Populations

The case of $\Sigma_1 \neq \Sigma_2 \neq \cdots \neq \Sigma_g$

An important special case occur when the

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} \mid \Sigma_i \mid^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu_i)'\Sigma_i^{-1}(\mathbf{x} - \mu_i)\right], \quad i = 1, 2, \cdots, g \tag{4.7}$$

are multivariate normal densities with mean vector μ_i and covariance matrices Σ_i. If further $c(i \mid i) = 0$, $c(k \mid i) = 1$, $k \neq i$, or, equivalently, the misclassification costs are equal, (4.5) becomes :

Allocate $\mathbf{x}$ to π_k if

$$\begin{aligned} \ln p_k f_k(\mathbf{x}) &= \ln p_k - \left(\frac{p}{2}\right)\ln(2\pi) - \frac{1}{2}\ln \mid \Sigma_k \mid - \frac{1}{2}(\mathbf{x} - \mu_k)'\Sigma_k^{-1}(\mathbf{x} - \mu_k) \\ &= \max_i \ln p_i f_i(\mathbf{x}) \end{aligned} \tag{4.8}$$

The constant $(p/2)\ln(2\pi)$ can be ignored in (4.8) since it is the same for all populations. We therefore define the ***quadratic discrimination score*** for the ith population to be

$$d_i^Q(\mathbf{x}) = -\frac{1}{2}\ln|\Sigma_i| - \frac{1}{2}(\mathbf{x}-\mu_i)'\Sigma_i^{-1}(\mathbf{x}-\mu_i) + \ln p_i, \quad i = 1,2,\cdots,g \tag{4.9}$$

The quadratic score, $d_i^Q(\mathbf{x})$, is composed of contributions from the generalized variance $|\Sigma_i|$, the prior probability p_i, and the squared distance from $\mathbf{x}$ to the population mean μ_i. Using discriminant scores the classification rule of (4.8) becomes the following.

Minimum Total Probability of Misclassification Rule for Normal Populations

Allocate $\mathbf{x}$ to π_k if

the quadratic score $d_k^Q(\mathbf{x})$ = largest of $d_1^Q(\mathbf{x}), d_2^Q(\mathbf{x}), \cdots, d_g^Q(\mathbf{x})$

(4.10)

where $d_i^Q(\mathbf{x})$ is given by (4.9), $i = 1,2,\cdots,g$.

In practice, the μ_i and Σ_i are unknown, but a training set of correctly classified observations is often available for the construction of estimates. The relevant sample quantities for population π_i are

$$\begin{aligned} \bar{\mathbf{x}}_i &= \text{sample mean vector} \\ \mathbf{S}_i &= \text{sample covariance matrix, and} \\ n_i &= \text{sample size} \end{aligned}$$

The estimate of the quadratic discriminant score $\hat{d}_i^Q(\mathbf{x})$ is then

$$\hat{d}_i^Q(\mathbf{x}) = -\frac{1}{2}\ln|\mathbf{S}_i| - \frac{1}{2}(\mathbf{x}-\bar{\mathbf{x}}_i)'\mathbf{S}_i^{-1}(\mathbf{x}-\bar{\mathbf{x}}_i) + \ln p_i \tag{4.11}$$

and the classification rule based on the sample is as follows.

Estimated Minimum Total Probability of Misclassification Rule for Normal Populations

Allocate $\mathbf{x}$ to π_k if

the quadratic score $\hat{d}_k^Q(\mathbf{x})$ = the largest of $\hat{d}_1^Q(\mathbf{x}), \hat{d}_2^Q(\mathbf{x}), \cdots, \hat{d}_g^Q(\mathbf{x})$

(4.12)

where $\hat{d}_i^Q(\mathbf{x})$ is given by (4.11), $i = 1,2,\cdots,g$.

The case of $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \cdots = \boldsymbol{\Sigma}_g = \boldsymbol{\Sigma}$

A simplification is possible if the population covariance matrices, $\boldsymbol{\Sigma}_i$, are equal. When $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}$, for $i = 1, 2, \cdots, g$, the discriminant score in (4.9) becomes

$$d_i^Q(\mathbf{x}) = -\frac{1}{2}\ln|\boldsymbol{\Sigma}| - \frac{1}{2}\mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{x} + \mu_i'\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\mu_i'\boldsymbol{\Sigma}^{-1}\mu_i + \ln p_i$$

The first two terms are the same for $d_1^Q(\mathbf{x}), d_2^Q(\mathbf{x}), \cdots, d_g^Q(\mathbf{x})$, and, consequently, they can be ignored for allocation purposes. The remaining terms consist of a constant $c_i = \ln p_i - \frac{1}{2}\mu_i'\boldsymbol{\Sigma}^{-1}\mu_i$ and a *linear* combination of the components of $\mathbf{x}$. Defining the ***linear discriminant score***

$$d_i(\mathbf{x}) = \mu_i'\boldsymbol{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}\mu_i'\boldsymbol{\Sigma}^{-1}\mu_i + \ln p_i \tag{4.13}$$

we obtain the following form of the allocation rule.

Minimum Total Probability of Misclassification Rule for Equal Covariance Normal Populations

Allocate $\mathbf{x}$ to π_k if

the linear discriminant score $d_k(\mathbf{x})$ = largest of $d_1(\mathbf{x}), d_2(\mathbf{x}), \cdots, d_g(\mathbf{x})$

(4.14)

where $d_i(\mathbf{x})$ is given by (4.13), $i = 1, 2, \cdots, g$.

The estimate $\hat{d}_i(\mathbf{x})$, of the linear discriminant score $d_i(\mathbf{x})$ is based on the pooled estimate of $\boldsymbol{\Sigma}$,

$$\mathbf{S}_{pooled} = \frac{(n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + \cdots + (n_g-1)\mathbf{S}_g}{n_1 + n_2 + \cdots + n_g - g} \tag{4.15}$$

and is given by

$$\hat{d}_i(\mathbf{x}) = \mathbf{x}_i'\mathbf{S}_{pooled}^{-1}\mathbf{x} - \frac{1}{2}x_i'\mathbf{S}_{pooled}^{-1}\mathbf{x}_i + \ln p_i \tag{4.16}$$

Consequently, we have the following :

Estimated Minimum Total Probability of Misclassification Rule for Equal Covariance Normal Populations

Allocate $\mathbf{x}$ to π_k if

the linear discriminant score $\hat{d}_k(\mathbf{x})$ = the largest of $\hat{d}_1(\mathbf{x}), \hat{d}_2(\mathbf{x}), \cdots, \hat{d}_g(\mathbf{x})$

(4.17)

with $\hat{d}_i(\mathbf{x})$ is given by (4.16), $i = 1, 2, \cdots, g$.

Special cases

1. Expressions (4.13) and (4.16) are convenient linear functions of $\mathbf{x}$. An equivalent classifier for the equal covariance case can be obtained from (4.9) by ignoring the constant term, $-\ln|\mathbf{\Sigma}|$. The result can then be interpreted in terms of the squared distances

$$D_i^2(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}}_i)'\mathbf{S}_{pooled}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i) \tag{4.18}$$

from $\mathbf{x}$ to the sample mean vector $\bar{\mathbf{x}}$. The allocatory rule is then

Assign $\mathbf{x}$ to the popuiation π_i for which $-\frac{1}{2}D_i^2(\mathbf{x}) + \ln p_i$ is largest

(4.19)

We see that this rule or, equivalently, (4.17) assign $\mathbf{x}$ to the "closest" population. (The distance measure is penalized by $\ln p_i$.)

If the prior probabilities are unknown, the usual procedure is to set $p_1 = p_2 = \cdots = p_g = 1/g$. An observation is then assigned to the closest population.

2. A second form of the classification rule in (4.14), obtained by comparing the $d_i(\mathbf{x})$ two at a time, merits consideration. The condition $d_k(\mathbf{x})$ is the largest linear discriminant score among $d_1(\mathbf{x}), d_2(\mathbf{x}), \cdots, d_g(\mathbf{x})$ is equivalent to

$$\begin{aligned} 0 &\leq d_k(\mathbf{x}) - d_i(\mathbf{x}) \\ &= (\mu_k - \mu_i)'\mathbf{\Sigma}^{-1}\mathbf{x} - \frac{1}{2}(\mu_k - \mu_i)'\mathbf{\Sigma}^{-1}(\mu_k + \mu_i) + \ln\left(\frac{p_k}{p_i}\right) \end{aligned}$$

for all $i = 1, 2, \cdots, g$.

Adding $-\ln(p_k/p_i) = \ln(p_i/p_k)$ to both sides of the inequality above gives the alternate form of the classification rule which minimizes the total probability of misclassification. Thus we :

Allocate $\mathbf{x}$ to π_k if

$$(\mu_k - \mu_i)'\Sigma^{-1}\mathbf{x} - \frac{1}{2}(\mu_k - \mu_i)'\Sigma^{-1}(\mu_k + \mu_i) \geq \ln\left(\frac{p_i}{p_k}\right) \tag{4.20}$$

for all $i = 1, 2, \cdots, g$.

Denote the left-hand side of (4.20) by $d_{ki}(\mathbf{x})$. The conditions in (4.20) define classification regions $R_1, R_2, \cdots, R_g$, which are separated by (hyper) planes. This follows because $d_{ki}(\mathbf{x})$ is a linear combination of the components of $\mathbf{x}$. For example, when $g = 3$, the classification region R_1 consists of all $\mathbf{x}$ satisfying

$$R_1 : d_{1i}(\mathbf{x}) \geq \ln\left(\frac{p_i}{p_1}\right) \text{ for } i = 2, 3$$

That is, R_1 consists of those $\mathbf{x}$ for which

$$d_{12}(\mathbf{x}) = (\mu_1 - \mu_2)'\Sigma^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 + \mu_2) \geq \ln\left(\frac{p_2}{p_1}\right)$$

and, *simultaneously*,

$$d_{13}(\mathbf{x}) = (\mu_1 - \mu_3)'\Sigma^{-1}\mathbf{x} - \frac{1}{2}(\mu_1 - \mu_3)'\Sigma^{-1}(\mu_1 + \mu_3) \geq \ln\left(\frac{p_3}{p_1}\right)$$

Assuming μ_1, μ_2, and μ_3 do not lie along a straight line, the equation $d_{12}(\mathbf{x}) = \ln(p_2/p_1)$ and $d_{13}(\mathbf{x}) = \ln(p_3/p_1)$ define two intersecting hyperplanes that delineate R_1 in the p-dimensional variable space. The term $\ln(p_2/p_1)$ places the plane closer to μ_1 than μ_2 if p_2 is greater than p_1.

Since (4.14) and (4.20) are just two equivalent forms of the minimum TPM rule, the classification regions are the same for both cases.

The sample version of the alternative form in (4.20) is obtained by substituting $\mathbf{x}_i$ for μ_i and inserting the pooled sample covariance matrix $\mathbf{S}_{pooled}$ for Σ. When $\sum_{i=1}^{g}(n_i - 1) > p$, so that $\mathbf{S}_{pooled}^{-1}$ exists, this sample analog becomes :

Allocate $\mathbf{x}$ to π_k if

$$\begin{aligned} \hat{d}_{ki}(\mathbf{x}) &= (\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_i)'\mathbf{S}_{pooled}^{-1}\mathbf{x} - \frac{1}{2}(\bar{\mathbf{x}}_k - \bar{\mathbf{x}}_i)'\mathbf{S}_{pooled}^{-1}(\bar{\mathbf{x}}_k + \bar{\mathbf{x}}_i) \\ &\geq \ln\left(\frac{p_i}{p_k}\right) \text{ for all } i \neq k \end{aligned} \tag{4.21}$$

Given the fixed training set values $\bar{\mathbf{x}}_i$ and $\mathbf{S}_{pooled}$, $\hat{d}_{ki}(\mathbf{x})$ is a linear function of the components of $\mathbf{x}$. Therefore, the classification regions defined by (4.21) - or, equivalently, by (4.17) - are also bounded by hyperplanes.

As with the sample linear discriminant rule of (4.17), if the prior probabilities are difficult to assess, they are frequently all taken to be equal. In this case, $\ln(p_i/p_k) = 0$ for all pairs.

4.2.3 General Evaluation

Because they employ estimates of population parameters, the sample classification rules of (4.12) and (4.17) may no longer be optimal. Their performance, however, can be evaluated using Lachenbruch's holdout procedure. If $n_{iM}^{(H)}$ is the number of misclassified holdout observations in the ith group, $i = 1, 2, \cdots, g$, then an estimate of the expected actual error rate, $E(\text{AER})$, is provided by

$$\hat{E}(\text{AER}) = \frac{\sum_{i=1}^{g} n_{iM}^{(H)}}{\sum_{i=1}^{g} n_i} \tag{4.22}$$

Our discussion has tended to emphasize the linear discriminant rule of (4.17) and (4.21), and many commercial computer programs are based upon it. Although the linear discriminant rule has a a simple structure, we must remember that it was derived under the strong assumptions of multivariate normality and equal covariances. Before implementing a linear classification rule, these tentative assumptions should be checked in the order : multivariate normality, then equality of covariances. If one or both of these assumptions is violated, improved classification is probably possible if the data are first transformed.

The quadratic rules are an alternative to classification with linear discriminant functions. They are appropriate if normality appears to hold but the assumption of equal covariance matrices is seriously violated. However, the assumption of normality seems to be more critical for quadratic rules than linear rules. If doubt exists as to the appropriateness of a linear or quadratic rule, both rules can be constructed and their error rates examined using Lachenbruch's holdout procedure.

4.2.4 Implementation consideration for Bayes Procedure

Most of discrimination and allocation rules in Bayes Procedure assume that the underlying populations have multivariate normal distribution. At the very least the marginal frequency distributions of the multivariate observations on each variable can be checked for normality. This ought to be done for the sample from every group. If the data are not multivariate normal, two options are available. First, the nonnormal data can be transformed to data more nearly normal. Second, one can use the rules without worrying about the form of the parent populations and hope that it will work reasonably well [Johnson, *et al*, 1988]. In this research, the second option will be chosen by taking a moderately large sample. It is hoped that if the sample size is large then the statistics from the sample become indistinguishable from the parameters of the normal population.

Discrimination and allocation rules are of two kinds

(a) linear rules when the covariance matrices of the groups are equal, and

(b) quadratic rules when they are not equal.

Algorithm (4.12) and formula (4.11), as well as the estimate of the expected actual error rate (4.22) are implemented by using programs written in Pascal.

4.3 Fisher's Method for Several Populations : Canonical Analysis

Fisher also proposed a several population extension of his discriminant method discussed in Section 3.3. The motivation behind the Fisher discriminant analysis is the need to obtain a reasonable representation of the population that involves only a *few* linear combinations of the observations, such as $\ell_1'\mathbf{x}$, $\ell_2'\mathbf{x}$, and $\ell_3'\mathbf{x}$. This approach is usually known as Canonical Analysis [James, 1985] and has several advantages when one is interested in *separating* several populations for (1) visual inspection or (2) graphical decriptive purposes. It allows for the following.

1. Convenient representations of the g populations that reduce the dimension from a very large number of characteristics to a relatively few linear combinations. Of course, some information needed for optimal classification – may be lost unless the population means lie completely in the lower-dimensional space selected.

2. Plotting of the means of the first two or three linear combinations (discriminants). This helps display the relationships and possible groupings of the populations.

3. Scatterplots of the sample values of the first two discriminants, which can indicate outliers or other abnormalities in the data.

The primary purpose of Fisher's discriminant analysis is to *separate* populations. It can, however, also be used to classify, and we shall indicate this use. It is not necessary to assume that the g populations are multivariate normal. However, we do assume the $p \times p$ population covariance matrices are equal and of full rank. [1] That is, $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_g = \Sigma$ [Johnson *et al*, 1988].

Let μ denote the mean vector of the combined groups and $\mathbf{B}_0$ the between groups sums of cross-products so that

$$\mathbf{B}_0 = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})', \quad \text{where} \quad \bar{\mu} = \frac{1}{g} \sum_{i=1}^{g} \mu_i \tag{4.23}$$

We consider the linear combination

$$Y = \ell' \mathbf{X}$$

which has expected value

$$E(Y) = \ell' E(\mathbf{X} \mid \pi_i) = \ell' \mu_i \quad \text{for population } \pi_i$$

and variance

$$\text{Var}(Y) = \ell' \text{Cov}(\mathbf{X}) \ell = \ell' \Sigma \ell \quad \text{for all populations}$$

Consequently, the expected value $\mu_{iY} = \ell' \mu_i$ changes as the population from which $\mathbf{X}$ is selected changes. We first define the overall mean

$$\mu_Y = \frac{1}{g} \sum_{i=1}^{g} \mu_{iY} = \frac{1}{g} \sum_{i=1}^{g} \ell' \mu_i$$

---

[1] If not, we let $P = [\mathbf{e}_1, \cdots, \mathbf{e}_q]$ be the eigenvectors of Σ corresponding to nonzero eigenvalues $[\lambda_1, \cdots, \lambda_q]$. Then we replace $\mathbf{X}$ by $\mathbf{P}'\mathbf{X}$, which has a full rank covariance matrix $\mathbf{P}'\Sigma\mathbf{P}$.

$$
\begin{aligned}
&= \ell'\left(\frac{1}{g}\sum_{i=1}^{g}\mu_i\right) \\
&= \ell'\mu
\end{aligned}
$$

and form the ratio

$$
\begin{aligned}
\frac{\left(\begin{array}{c}\text{Sum of squared distance from}\\ \text{population to overall mean of } Y\end{array}\right)}{(\text{Variance of } Y)} &= \frac{\sum_{i=1}^{g}(\mu_{iY}-\mu_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^{g}(\ell'\mu_i-\ell'\mu)^2}{\ell'\Sigma\ell} \\
&= \frac{\ell'\left(\sum_{i=1}^{g}(\mu_i-\mu)(\mu_i-\mu)'\right)\ell}{\ell'\Sigma\ell}
\end{aligned}
$$

or

$$
\frac{\sum_{i=1}^{g}(\mu_{iY}-\mu_Y)^2}{\sigma_Y^2} = \frac{\ell'\mathbf{B}_0\ell}{\ell'\Sigma\ell} \tag{4.24}
$$

The ratio in (4.24), a multigroup generalization of (3.6), measures the variability *between* the groups of Y-values relative to the common variability *within* groups. Analogous to the two-population case, we can select ℓ to maximize the ratio of (4.24).

The result can be stated as the following.

Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_s > 0$ denote the $s \leq \min(g-1, p)$ nonzero eigenvalues of $\Sigma^{-1}\mathbf{B}_0$ and $\mathbf{e}_1, \mathbf{e}_2, \cdots, \mathbf{e}_s$ the corresponding eigenvectors (scaled so that $\mathbf{e}'\Sigma\mathbf{e} = 1$). Then the vector of coefficients ℓ that maximizes the ratio

$$
\frac{\ell'\mathbf{B}_0\ell}{\ell'\Sigma\ell} = \frac{\ell'\left[\sum_{i=1}^{g}(\mu_i-\mu)(\mu_i-\mu)'\right]\ell}{\ell'\Sigma\ell}
$$

is given by $\ell_1 = \mathbf{e}_1$. The linear combination $\ell_1'\mathbf{X}$ is called the ***first discriminant***. The value $\ell_2 = \mathbf{e}_2$ maximizes the ratio subject to $\text{Cov}(\ell_1'\mathbf{X}, \ell_2'\mathbf{X}) = 0$. The linear combination $\ell_2'\mathbf{X}$ is called the ***second discriminant***. Continuing, $\ell_k' = \mathbf{e}_k$ maximizes the ratio subject to $0 = \text{Cov}(\ell_k'\mathbf{X}, \ell_i'\mathbf{X})$, $i < k$, and $\ell_k'\mathbf{X}$ is called the kth discriminant. Also $\text{Var}(\ell_i'\mathbf{X}) = 1$, $i = 1, \cdots, s$.

Ordinarily, Σ and the μ_i are unavailable, but we have a training set consisting of correctly classified observations. Suppose the training set consists of a random sample of size n_i from population π_i, $i = 1, 2, \cdots, g$. Denote the $p \times n_i$ data sets, from population π_i, by $\mathbf{X}_i$ and its jth column by $\mathbf{x}_{ij}$. After first constructing the sample mean vectors

$$
\mathbf{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}\mathbf{x}_{ij}
$$

and the covariance matrices $\mathbf{S}_i$, $i = 1, 2, \cdots, g$, we define the "overall average" vector

$$\bar{\mathbf{x}} = \frac{\sum_{i=1}^{g} n_i \bar{\mathbf{x}}_i}{\sum_{i=1}^{g} n_i} = \frac{\sum_{i=1}^{g} \sum_{j=1}^{n_i} \mathbf{x}_{ij}}{\sum_{i=1}^{g} n_i}$$

which is the $p \times 1$ vector average taken over *all* of the sample observations in the training set.

Corresponding to the population between groups matrix $\mathbf{B}_0$ in (4.23), we define the *sample between group* matrix

$$\hat{\mathbf{B}}_0 = \sum_{i=1}^{g} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' \tag{4.25}$$

Also, an estimate of $\boldsymbol{\Sigma}$ is based on the *sample within groups* matrix.

$$\mathbf{W} = \sum_{i=1}^{g} (n_i - 1)\mathbf{S}_i = \sum_{i=1}^{g} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' \tag{4.26}$$

Consequently, $\mathbf{W}/(n_1 + n_2 + \cdots + n_g - g) = \mathbf{S}_{pooled}$ is the estimate of $\boldsymbol{\Sigma}$.

Before presenting the sample discriminants, we note that $\mathbf{W}$ is the constant $(n_1 + n_2 + \cdots + n_g - g)$ times $\mathbf{S}_{pooled}$, so the same $\hat{\ell}$ that maximizes $\hat{\ell}'\hat{\mathbf{B}}_0\hat{\ell}/\hat{\ell}'\mathbf{S}_{pooled}\hat{\ell}$ also maximizes $\hat{\ell}'\hat{\mathbf{B}}_0\hat{\ell}/\hat{\ell}'\mathbf{W}\hat{\ell}$. Moreover, we can present the optimizing $\hat{\ell}$ in the more customary form as eigenvectors, $\hat{\mathbf{e}}_i$, of $\mathbf{W}^{-1}\hat{\mathbf{B}}_0$, because if $\mathbf{W}^{-1}\hat{\mathbf{B}}_0\hat{\mathbf{e}} = \hat{\lambda}\hat{\mathbf{e}}$ then $\mathbf{S}_{pooled}^{-1}\hat{\mathbf{B}}_0\hat{\mathbf{e}} = \hat{\lambda}(n_1 + n_2 + \cdots + n_g - g)\hat{\mathbf{e}}$.

Fisher's Sample Discriminants

Let $\hat{\lambda}_1, \hat{\lambda}_2, \cdots, \hat{\lambda}_s > 0$ denote the $s \leq \min(g-1, p)$ nonzero eigenvalues of $\mathbf{W}^{-1}\hat{\mathbf{B}}_0$ and $\hat{\mathbf{e}}_1, \cdots, \hat{\mathbf{e}}_s$ be the corresponding eigenvectors (scaled so that $\hat{\mathbf{e}}'\mathbf{S}_{pooled}\hat{\mathbf{e}} = 1$). Then the vector of coefficients $\hat{\ell}$ that maximizes the ratio

$$\frac{\hat{\ell}'\hat{\mathbf{B}}_0\hat{\ell}}{\hat{\ell}'\mathbf{W}\hat{\ell}} = \frac{\hat{\ell}'\left(\sum_{i=1}^{g}(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})'\right)\hat{\ell}}{\hat{\ell}'\left[\sum_{i=1}^{g}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)'\right]\hat{\ell}} \tag{4.27}$$

is given by $\hat{\ell}_1 = \hat{\mathbf{e}}_1$. The linear combination $\hat{\ell}_1'\mathbf{x}$ is called the ***sample first discriminant.*** The choice $\hat{\ell}_2 = \hat{\mathbf{e}}_2$ produces the ***sample second discriminant***, $\hat{\ell}_2'\mathbf{x}$. Continuing, $\hat{\ell}_k'\mathbf{x} = \hat{\mathbf{e}}_k'\mathbf{x}$ is the ***sample kth discriminant***, $k \leq s$.

Unlike the population result, the discriminants will not have zero covariance for each random sample $\mathbf{X}_i$. Rather, the condition

$$\hat{\ell}_i'\mathbf{S}_{pooled}\hat{\ell}_k = \begin{cases} 1 & \text{if } i = k \leq s \\ 0 & \text{otherwise} \end{cases} \tag{4.28}$$

will be satisfied. The use of $\mathbf{S}_{pooled}$ is appropriate because we tentatively assumed that the g population covariance matrices were equal [Johnson *et al*, 1988].

4.3.1 Using Fisher's Discriminants to Classify

Fisher's discriminants were derived for the purpose of obtaining a low-dimensional representation of the data that separates the populations as much as possible. Although they were derived from separatory considerations, the discriminants also provide the basis for a classification rule.

Setting

$$Y_k = \ell_k'\mathbf{X} = k\text{th } discriminant,\ k \leq s \tag{4.29}$$

we conclude that

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix} \text{ has mean vector } \mu_{iY} = \begin{bmatrix} \mu_{iY_1} \\ \vdots \\ \mu_{iY_s} \end{bmatrix} = \begin{bmatrix} \ell_1'\mu_i \\ \vdots \\ \ell_s'\mu_i \end{bmatrix}$$

under population π_1 and covariance matrix $\mathbf{I}$, for all populations.

Because the components of $\mathbf{Y}$ have unit variances and zero covariances, the appropriate measure of squared distance from $\mathbf{Y} = \mathbf{y}$ to μ_{iY} is

$$(\mathbf{y} - \mu_{iY})'(\mathbf{y} - \mu_{iY}) = \sum_{j=1}^{s}(y_j - \mu_{iY_j})^2$$

A reasonable classification rule is one that assign $\mathbf{y}$ to population π_k if the squared distance from $\mathbf{y}$ to μ_{kY} is smaller than the squared distance from $\mathbf{y}$ to μ_{iY} for $i \neq k$.

If only r of the discriminants are used for allocation, the rule is :

Fisher's Classification Procedure Based on Sample Discriminants

Allocate $\mathbf{x}$ to π_k if

$$\begin{aligned} \sum_{j=1}^{r}(\hat{y}_j - y_{kj})^2 &= \sum_{j=1}^{r}[\hat{\ell}_j'(\mathbf{x} - \mathbf{x}_k)]^2 \\ &\leq \sum_{j=1}^{r}[\hat{\ell}_j'(\mathbf{x} - \mathbf{x}_i)]^2 \quad \text{for all } i \neq k \end{aligned} \tag{4.30}$$

where $\hat{\ell}_j$ is defined in (4.27) and $r \leq s$.

[Johnson *et al*, 1988].

4.4 Stepwise Discriminant Analysis

Even if, using canonical analysis, we succeed in reducing the number of dimensions used in classification, we still have to measure all of the original variables because each of the new dimensions is a linear combination of all of the old variables. [James, 1985]

In the previous section we considered the case of classifying a p-dimensional observation vector $\mathbf{x} = (x_1, \cdots, x_p)'$ into one of g multivariate normal populations $\pi_i : (\mu_i^{p\times 1}, \Sigma^{p\times p})$, where $\mu_i = (\mu_{i1}, \cdots, \mu_{ip})', i = 1, \cdots, g$. Since $\mathbf{x}$ is a realization of a random vector $\mathbf{X} = (X_1, \cdots, X_p)$, the results presented used all p variables $X_1, \cdots, X_p$ to discriminate between the g populations. In many applications, however, it is desired to identify a subset of these variables which "best" discriminates between the g populations. In this case an F statistic based on a one-way analysis of variance test is used to choose variables. The F statistic is called the *F-to-enter* for variables not chosen and the *F-to-remove* for chosen variables [Afifi *et al*, 1979].

For two populations, Lachenbruch [1975] mentioned a method proposed by Rao [1970] to test whether $(p-1)$ variables will do as good a job as p variables. The following statistic F is used as criterion.

$$F = n_1 + n_2 - p - 1\frac{C(D_p^2 - D_{p-1}^2)}{1 + CD_{p-1}^2}$$

where D_p^2 and D_{p-1}^2 are the Mahalanobis D^2 statistics on the full set and the subset, respectively (see formula 3.21 about Mahalanobis D^2), and

$$C = \frac{n_1 n_2}{(n_1 + n_2)(n_1 + n_2 - 2)}$$

This has an F distribution with 1 and $n_1 + n_2 - p - 1$.

This is only appropriate if one variable is to be eliminated. If a second variable is to be eliminated, the procedure should be repeated with the $p - 1$ remaining variables.

4.4.1 Analysis for multiple populations

In brief, the logic behind the stepwise discriminant procedure is as follows.

We first identify the variable for which the mean values in the g populations are "most different". For each variable this difference is measured by a one-way analysis of variance

F statistic, and the variable with the largest F is chosen (or *entered*). On successive steps, we consider the conditional distribution of each variable not entered given the variables entered. Of the variables not entered, we identify the variable for which the mean values of the conditional distributions in the g populations are "most different". This difference is also measured by a one-way analysis of variance F statistic. The stepwise process is stopped when no additional variables significantly contribute to the discrimination between the g populations. A value of minimum F-to-enter and a value of minimum F-to-remove less than the F-to-enter must be specified in advance.[Afifi *et al*, 1979]

4.4.2 The stepwise algorithm

In detail, an algorithm for the stepwise procedure is as follows [Afifi *et al*, 1979] :

Let $\mathbf{x}_{i1}^{p\times 1}, \cdots, \mathbf{x}_{in_i}^{p\times 1}$ be a random sample from $\pi_i, i = 1, \cdots, g$. Then, we have the following steps.

Step 0 The *F-to-enter* with $g - 1$ and $n - g$ degrees of freedom is computed for each $X_j, j = 1, \cdots, p$. This F-to-enter is the one-way analysis of variance F statistic for testing $H_0 : \mu_{1j} = \cdots = \mu_{gj}$, for $j = 1, \cdots, p$. If all F-to-enter are less than the minimum F-to-enter (specified in advance), the process is terminated, and we conclude that no variable significantly discriminates between the populations.

Step 1 The variable X_{j_1} having the largest F-to-enter is chosen as the first variable. The estimated linear discriminant coefficient and constant are calculated for each population $\pi_i, i = 1, \cdots, g$. The classification table and *F-approximation* statistic are also calculated. Also, the *F-to-remove*, with $g - 1$ and $n - g$ degrees of freedom, which is equal to the F-to-enter is calculated for X_{j_1}. Then, the F-to-enter with $g - 1$ and $n - g - 1$ degrees of freedom is calculated for each variable not entered. This tests the hypothesis $H_0 : \mu_{1j.j_1} = \cdots = \mu_{gj.j_1}$, where $\mu_{1j.j_1}$ is the mean of the conditional distribution in π_i of X_j given $X_{j_1}, i = 1, \cdots, g$, $j = 1, \cdots, p$, $j \neq j_1$. If all the F-to-enter are less than the minimum F-to-enter, then Step S is executed ; otherwise, Step 2 is executed.

Step 2 The variable X_{j_2} is chosen for which F-to-enter is maximum. The two estimated linear discriminant coefficients and constant are calculated for each population $\pi_i, i = 1, \cdots, g$.

The application of these coefficients and constants to the measurement vectors in the training sample produces the classification matrix given in Table 5.2; and the application to the validating sample, which consists of 2726 measurement vectors that are not involved in the computation of the coefficients and the constants, results in the classification matrix given in Table 5.3. The diagonal elements of a classification matrix represent the

Table 5.2: Classification matrix for the training sample : monoalphabetic 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.63	31	0	0	0	2	0	0	0	5	0	0	0	0	5
B	0.00	0	0	0	0	0	5	48	0	0	0	0	0	0	0
C	50.50	0	0	51	7	0	0	0	9	0	0	0	17	5	0
D	20.21	0	0	24	19	0	0	0	15	0	0	0	22	8	0
E	92.66	0	0	0	0	101	0	0	0	0	0	0	0	0	0
F	34.48	0	0	0	0	0	30	20	1	0	0	0	0	3	0
G	49.48	0	0	0	0	0	23	48	1	0	0	0	0	1	0
H	19.39	0	0	20	13	0	2	1	19	0	0	0	11	14	0
I	11.36	8	0	0	0	0	0	0	0	10	0	0	0	0	9
J	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
K	13.19	0	0	0	0	0	0	1	0	0	0	12	0	0	0
L	24.24	0	0	33	19	0	0	0	14	0	0	0	24	2	0
M	12.77	0	0	11	6	0	7	1	17	0	0	0	5	12	0
N	11.43	19	0	0	0	0	0	0	0	8	0	0	0	0	12
O	19.05	26	0	1	0	0	0	0	0	9	0	0	0	0	10
P	5.88	0	0	3	5	0	5	5	13	0	0	0	3	22	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
R	62.63	0	0	5	0	0	0	0	0	9	0	0	0	0	7
S	11.88	12	0	4	0	0	0	0	0	12	0	0	0	0	8
T	52.75	22	0	0	0	9	0	0	0	3	0	0	0	0	2
U	30.63	0	0	1	3	0	13	6	15	0	0	0	1	24	0
V	30.77	0	0	0	0	0	2	19	0	0	0	1	0	0	0
W	4.30	0	0	0	0	0	6	40	0	0	0	0	0	1	0
X	0.00	0	0	0	0	0	0	1	0	0	0	63	0	0	0
Y	28.71	0	0	0	0	0	12	49	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	101	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	26.81	118	0	153	72	112	105	239	104	56	0	369	83	92	53

number of known measurement vectors which are correctly classified into the actual groups. In Table 5.2, for example, 31 out of 98 measurement vectors for character group A are correctly classified into group A ; 51 out of 96 vectors for group C are correctly classified into group C. Thus, 31.63 percent of vectors for group A and 50.50 percent of vector for group B are correctly classified, and so on. The total percentage of correctness is calculated by adding the values of all diagonal elements and dividing this result to the total number

For each step the quantities that should be given are the step number, the variable entered or removed, the F-to-enter or F-to-remove, and the *F-approximation* statistic.

Jennrich *et al* [1977] proposed another algorithm for stepwise discriminant analysis based on Wilks' Λ-criterion. This algorithm is implemented by the subprogram BMDP7M of software package BMDP [Dixon *et al*, 1988].

4.4.3 Implementation consideration for Fisher's Method and Stepwise Procedure

The assumption in Fisher's method is that the population covariance matrices are equal. This means that this method is linear in nature. James [1985] mentioned that pooled sample covariance matrix can be considered as "the average" of different covariance matrices of all groups. In this research the implementation of Fisher's Canonical Analysis and Stepwise Linear Discriminant Analysis is done with the help of a commercial computer package called BMDP [Dixon *et al*, 1988]. The program in the package related to the analysis is BMDP7M.

4.5 Implementation

The purpose of this section is to elaborate the implementation of the procedures and methods mentioned in the previous sections.

A sample of three hundred English text files is used in the processes to generate various variables in relation to monoalphabetic and polyalphabetic substitutions.

The first one hundred text files are used to estimate prior probability of every character group in population of English texts. The relative frequencies of character groups in a text are calculated. The average value of this variable for a character group – which is calculated from one hundred text files – is used as the estimate of the prior probability of the character group. The following table (Table 4.1) gives the prior probability estimates of character groups in monoalphabetic and polyalphabetic substitutions.

The remaining two hundred text files are used to calculate two hundred *measurement vectors* for every character group. Each measurement vector consists of the values of variables which will be involved in the formation of discriminant functions for every character. One additional variable is inserted into the vector for the purpose of identifying particular

Table 4.1: Prior Probability Estimate of Every Character in Source Text

Group of Characters	Prior probabilities			
	Monoalphabetic Substitution	Polyalphabetic Substitution		
		Position 1	Position 2	Position 3
A	0.071292	0.071416	0.071679	0.070839
B	0.011646	0.011408	0.011962	0.011566
C	0.034395	0.034685	0.033948	0.034582
D	0.031729	0.032001	0.031707	0.031424
E	0.096478	0.097194	0.095476	0.096790
F	0.017467	0.017615	0.017284	0.017518
G	0.016217	0.015867	0.016488	0.016284
H	0.028782	0.029562	0.028622	0.028188
I	0.062482	0.062565	0.063247	0.061692
J	0.001170	0.001053	0.001272	0.001186
K	0.005394	0.005090	0.005156	0.005941
L	0.033322	0.033115	0.032920	0.033960
M	0.026321	0.026608	0.026556	0.025823
N	0.063462	0.063871	0.063108	0.063465
O	0.064976	0.067077	0.064464	0.063446
P	0.023460	0.023251	0.023090	0.024049
Q	0.001143	0.001162	0.001123	0.001146
R	0.054609	0.053476	0.055372	0.055002
S	0.059985	0.060728	0.060356	0.058841
T	0.074318	0.072276	0.073915	0.076833
U	0.023310	0.023237	0.023521	0.023193
V	0.009304	0.009434	0.009207	0.009280
W	0.012535	0.012519	0.012350	0.012746
X	0.002715	0.002665	0.002767	0.002716
Y	0.013247	0.013197	0.012798	0.013758
Z	0.000875	0.000703	0.000855	0.001067
@	0.159366	0.158226	0.160756	0.158665

character groups. The two hundred text files will give $200 \times 27 = 5400$ measurement vectors which are then saved in a data file for further processing.

In monoalphabetic substitution, one measurement vector of a character will have twelve variable values and one additional variable value to identify the character. The names of the variables are :

```
[ RELFREQ , ROWCOLSS, WINROWSS, WINCOLSS, RWCPROSS, CWCPROSS,
  INFOCONT, CTOENTRO, RCTOJENT, CCTOJENT, RWCENTRO, CWCENTRO,
  and INDICE as the index of character group (INDICE = 1,..27)].
```

The name of the data file to save the measurement vectors is *RESULT01.001*.

In polyalphabetic substitution, the measurement vector will have twenty variable values plus one variable value for indexing the characters. The names of the variables in the measurement vector are :

```
[ RELFREQ , ROWCOLSS, WINROWS3, WINCOLS3, RWCPROS3, CWCPROS3,
  INFOCONT, CTOENTRO, RCTOJEN3, CCTOJEN3, RWCENTR3, CWCENTR3,
  WINROWSS, RWCPROSS, RCTOJENT, RWCENTRO, WINCOLSS, CWCPROSS,
  CCTOJENT, CWCENTRO, and INDICE]
```

The names of data files to save the measurement vectors in polyalphabetic substitution with keylength = 3 are :

(a) *RESULT03.001* for Position 1,

(b) *RESULT03.002* for Position 2, and

(c) *RESULT03.003* for Position 3.

About fifty percent of measurement vectors of character groups are used to build the discriminant functions, both linear and quadratic. The same measurement vectors are used as "training samples" to determine the power of the discriminant functions in identifying character groups based on the values of the variables in their measurement vectors. The remaining fifty percent of measurement vectors - which are not involved in the formation of discriminant functions - are used as "validating samples" with the same purpose, *i.e.*, to determine the discriminating power of the functions.

As mentioned in the previous chapter, this research will compare the discriminant functions - both linear and quadratic - built by using

```
---------------------------------------------------------------------
/INPUT       FILE IS 'RESULT01.001'.
             VARIABLES = 13.
             FORMAT = '4F12.6/4F12.6/4F12.6,F7.0'.
/VARIABLES   NAMES ARE RELFREQ,  ROWCOLSS, WINROWSS,
                       WINCOLSS, RWCPROSS, CWCPROSS,
                       INFOCONT, CTOENTRO, RCTOJENT,
                       CCTOJENT, RWCENTRO, CWCENTRO, INDICE.
             GROUPING = INDICE.
/TRANSFORM   IF(RNDU(7832) LE .5) THEN INDICE=INDICE+27 .
/GROUP       CODES(INDICE) = 1 TO 54.
             NAMES(INDICE) = A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,
                             U,V,W,X,Y,Z,SP,
                             A1,B1,C1,D1,E1,F1,G1,H1,I1,J1,K1,L1,M1,N1,
                             O1,P1,Q1,R1,S1,T1,U1,V1,W1,X1,Y1,Z1,SP1.
             USE = A TO SP.
    PRIOR =  0.071292,0.011646,0.034395,0.031729,0.096478,0.017467,
             0.016217,0.028782,0.062482,0.001170,0.005394,0.033322,
             0.026321,0.063462,0.064976,0.023460,0.001143,0.054609,
             0.059985,0.074318,0.023310,0.009304,0.012535,0.002715,
             0.013247,0.000875,0.159366.
/PRINT       LINE=120. NO STEP. NO POST. NO POINT.
/PLOT        NO CANON.
/END
---------------------------------------------------------------------
```

Figure 4.1: The example of BMDP7M commands for stepwise linear discriminant analysis of monoalphabetic substitution by using twelve variables

(a) one variable [RELFREQ],

(b) three variables generated from one-dimensional probability space [RELFREQ, INFOCONT, CTOENTRO], and

(c) twelve and twenty variables in the case of monoalphabetic and polyalphabetic substitution, respectively.

Stepwise linear discriminant analysis is accomplished with the help of a commercial computer package BMDP [Dixon, 1988], and the name of the program in this package used for the analysis is BMDP7M. The example of commands for using BMDP7M in the analysis for monoalphabetic substitution is given in Figure 4.1.

For analysis using three variables, the following commands in Figure 4.1 :

```
            VARIABLES = 13.
            FORMAT = '4F12.6/4F12.6/4F12.6,F7.0'.
/VARIABLES  NAMES ARE RELFREQ,  ROWCOLSS, WINROWSS,
                      WINCOLSS, RWCPROSS, CWCPROSS,
                      INFOCONT, CTOENTRO, RCTOJENT,
                      CCTOJENT, RWCENTRO, CWCENTRO, INDICE.
```

are changed into :

```
            VARIABLES = 4.
            FORMAT = 'F12.6/24X,2F12,6/48X,F7.0'.
/VARIABLES  NAMES ARE RELFREQ , INFOCONT , CTOENTRO, INDICE.
```

and for analysis of one variable the commands become

```
            VARIABLES = 2.
            FORMAT = 'F12.6//48X,F7.0'.
/VARIABLES  NAMES ARE RELFREQ , INDICE.
```

The example of commands for using BMDP7M in the analysis for polyalphabetic substitution is shown in Figure 4.2 The similar modifications as in monoalphabetic substitution are done in analysis using one variable and three variables.

Computer programs in PASCAL language for accomplishing Quadratic Discriminant Analysis have been written and applied in this research for one variable, three variables, and twelve variables (monoalphabetic) or twenty variables (polyalphabetic).

The next chapter presents the results of the computations.

```
--------------------------------------------------------------------
/INPUT      FILE IS 'RESULTC3.003'.
            VARIABLES = 21.
            FORMAT = '4F12.6/4F12.6/4F12.6/4F12.6/4F12.6,F7.0'.
/VARIABLES  NAMES ARE RELFREQ,  ROWCOLSS, WINROWS3, WINCOLS3,
            RWCPROS3, CWCPROS3, INFOCONT, CTOENTRO, RCTOJEN3,
            CCTOJEN3, RWCENTR3, CWCENTR3, WINROWSS, RWCPROSS,
            RCTOJENT, RWCENTRO, WINCOLSS, CWCPROSS, CCTOJENT,
            CWCENTRO, INDICE.
            GROUPING = INDICE.
/TRANSFORM  IF(RNDU(7832) LE .5) THEN INDICE=INDICE+27 .
/GROUP      CODES(INDICE) = 1 TO 54.
            NAMES(INDICE) = A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,
                            U,V,W,X,Y,Z,SP,
                            A1,B1,C1,D1,E1,F1,G1,H1,I1,J1,K1,L1,M1,N1,
                            O1,P1,Q1,R1,S1,T1,U1,V1,W1,X1,Y1,Z1,SP1.
            USE = A TO SP.
 PRIOR=0.070839,0.011566,0.034582,0.031424,0.096790,0.017518,0.016284,
       0.028188,0.061692,0.001186,0.005941,0.033960,0.025823,0.063465,
       0.063446,0.024049,0.001146,0.055002,0.058841,0.076833,0.023193,
       0.009280,0.012746,0.002716,0.013758,0.001067,0.158665.
/PRINT      LINE=120. NO STEP. NO POST. NO POINT.
/PLOT       NO CANON.
/END
--------------------------------------------------------------------
```

Figure 4.2: The example of BMDP7M commands for stepwise linear discriminant analysis of polyalphabetic substitution, position 3 by using twenty variables

CHAPTER 5

COMPUTATIONS AND RESULTS

5.1 Introduction

This chapter presents the computational results from the linear and quadratic methods of discriminant analysis for monoalphabetic and polyalphabetic substitutions described in Chapter 4. Appendix A gives the details of the results from linear discriminant analysis, and Appendix B shows the results from quadratic discriminant analysis.

The results from stepwise linear discriminant analysis are the sets of coefficients and the constants of linear discriminant functions, and the results from quadratic discriminant analysis are quadratic discriminant functions, in each case, one function for every group. The term "group" is used here to denote the set of all occurences of a character type in a sample of text. The term "measurement vector" is a set of measurements made on each group.

Each of these functions is then used to calculate the *discriminant score* of each group, given the values of variables in a measurement vector. Thus, if we have twenty seven groups then we will have twenty seven discriminant scores. The maximum of these scores suggests that the measurement vector belongs to the group which has that maximum. The application of these functions to the known measurement vectors in the "training sample" or the "validating sample" will enable us to evaluate the discriminating power of the particular method by observing the *classification matrix*. This matrix shows the number of measurement vectors which are correctly classified, and the number which are incorrectly classified.

Table 5.1: Coefficient of linear discrimination for character groups, monoalphabetic, 1 variable

Group	Coefficients of RELFREQ	CONSTANT
A	2451.34766	-88.34501
B	427.62061	-7.06080
C	1247.35107	-25.56044
D	1107.85669	-20.95539
E	3437.08569	-170.82770
F	642.20062	-9.92955
G	527.38367	-8.08854
H	1000.28503	-17.81850
I	2162.29590	-69.45689
J	37.09273	-6.77037
K	187.06538	-5.72156
L	1156.78723	-22.48681
M	931.46771	-16.01187
N	2227.84033	-73.54531
O	2281.71167	-76.98656
P	835.21796	-13.70173
Q	47.97626	-6.80693
R	1933.44788	-56.22342
S	2099.72778	-65.69437
T	2612.28491	-99.92624
U	807.90601	-13.06809
V	322.35156	-6.15932
W	427.76773	-6.98903
X	109.97643	-6.08146
Y	442.89038	-7.12157
Z	23.69461	-7.04929
@	5575.79346	-445.24588

As an example, Table 5.1 shows the coefficients and the constants of linear discriminant functions for the twenty seven groups of characters [A,B,...,Z,@]. These functions are the results of the linear discriminant analysis using a training sample of 2674 measurement vectors, which consist of one variable [RELFREQ]. The coefficients and the constants can be used to classify a measurement vector which consists of only one value of variable RELFREQ into one of the character groups. The value of RELFREQ in the measurement vector is multiplied by the coefficient and is then added to the constant of every group. In this way, we will obtain twenty seven scores, one for each character type. The maximum of these scores indicates that the measurement vector belongs to the character type that has the maximum score.

The application of these coefficients and constants to the measurement vectors in the training sample produces the classification matrix given in Table 5.2; and the application to the validating sample, which consists of 2726 measurement vectors that are not involved in the computation of the coefficients and the constants, results in the classification matrix given in Table 5.3. The diagonal elements of a classification matrix represent the

Table 5.2: Classification matrix for the training sample : monoalphabetic 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.63	31	0	0	0	2	0	0	0	5	0	0	0	0	5
B	0.00	0	0	0	0	0	5	48	0	0	0	0	0	0	0
C	50.50	0	0	51	7	0	0	0	9	0	0	0	17	5	0
D	20.21	0	0	24	19	0	0	0	15	0	0	0	22	8	0
E	92.66	0	0	0	0	101	0	0	0	0	0	0	0	0	0
F	34.48	0	0	0	0	0	30	20	1	0	0	0	0	3	0
G	49.48	0	0	0	0	0	23	48	1	0	0	0	0	1	0
H	19.39	0	0	20	13	0	2	1	19	0	0	0	11	14	0
I	11.36	8	0	0	0	0	0	0	0	10	0	0	0	0	9
J	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
K	13.19	0	0	0	0	0	0	1	0	0	0	12	0	0	0
L	24.24	0	0	33	19	0	0	0	14	0	0	0	24	2	0
M	12.77	0	0	11	6	0	7	1	17	0	0	0	5	12	0
N	11.43	19	0	0	0	0	0	0	0	8	0	0	0	0	12
O	19.05	26	0	1	0	0	0	0	0	9	0	0	0	0	10
P	5.88	0	0	3	5	0	5	5	13	0	0	0	3	22	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
R	62.63	0	0	5	0	0	0	0	0	9	0	0	0	0	7
S	11.88	12	0	4	0	0	0	0	0	12	0	0	0	0	8
T	52.75	22	0	0	0	9	0	0	0	3	0	0	0	0	2
U	30.63	0	0	1	3	0	13	6	15	0	0	0	1	24	0
V	30.77	0	0	0	0	0	2	19	0	0	0	1	0	0	0
W	4.30	0	0	0	0	0	6	40	0	0	0	0	0	1	0
X	0.00	0	0	0	0	0	0	1	0	0	0	63	0	0	0
Y	28.71	0	0	0	0	0	12	49	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	101	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	26.81	118	0	153	72	112	105	239	104	56	0	369	83	92	53

number of known measurement vectors which are correctly classified into the actual groups. In Table 5.2, for example, 31 out of 98 measurement vectors for character group A are correctly classified into group A ; 51 out of 96 vectors for group C are correctly classified into group C. Thus, 31.63 percent of vectors for group A and 50.50 percent of vector for group B are correctly classified, and so on. The total percentage of correctness is calculated by adding the values of all diagonal elements and dividing this result to the total number

Table 5.2 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	31.63	19	0	0	1	5	30	0	0	0	0	0	0	0	98
B	0.00	6	0	0	0	0	0	2	8	2	0	31	0	0	96
C	50.50	0	2	0	9	0	0	1	0	0	0	0	0	0	101
D	20.21	0	1	0	2	0	0	3	0	0	0	0	0	0	94
E	92.66	0	0	0	0	0	8	0	0	0	0	0	0	0	109
F	34.48	0	1	0	0	0	0	31	0	0	0	1	0	0	87
G	49.48	6	0	0	0	0	0	10	2	0	0	12	0	0	97
H	19.39	0	3	0	0	0	0	14	0	1	0	0	0	0	98
I	11.36	16	0	0	20	17	8	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	4	0	0	0	0	0	100
K	13.19	0	0	0	0	0	0	1	69	3	0	5	0	0	91
L	21.24	0	1	0	4	0	0	2	0	0	0	0	0	0	99
M	12.77	0	11	0	2	0	0	22	0	0	0	0	0	0	94
N	11.43	19	0	0	19	17	11	0	0	0	0	0	0	0	105
O	19.05	20	0	0	12	11	16	0	0	0	0	0	0	0	105
P	5.88	0	6	0	0	0	0	40	0	0	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	6	0	0	0	0	0	102
R	62.63	3	0	0	62	12	1	0	0	0	0	0	0	0	99
S	11.88	11	0	0	34	12	8	0	0	0	0	0	0	0	101
T	52.75	4	0	0	2	1	48	0	0	0	0	0	0	0	91
U	30.63	0	14	0	0	0	0	34	0	0	0	0	0	0	111
V	30.77	0	0	0	0	0	0	0	32	14	0	36	0	0	104
W	4.30	0	0	0	0	0	0	3	7	4	0	32	0	0	93
X	0.00	0	0	0	0	0	0	0	40	1	0	0	0	0	105
Y	28.71	0	0	0	0	0	0	2	3	6	0	29	0	0	101
Z	0.00	0	0	0	0	0	0	0	1	0	0	0	0	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
Total	26.81	92	39	0	167	75	130	165	172	31	0	146	0	101	2674

of measurement vectors (see the summary of results for the training and validating samples in Table 5.4).

Table 5.3: Classification matrix for the validating sample : monoalphabetic 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.37	32	0	0	0	1	0	0	0	6	0	0	0	0	5
B	0.00	0	0	0	0	0	6	40	0	0	0	0	0	1	0
C	47.47	0	0	47	11	0	0	0	10	0	0	0	13	9	0
D	13.21	0	0	37	14	0	0	0	20	0	0	0	20	9	0
E	93.41	0	0	0	0	85	0	0	0	0	0	0	0	0	0
F	33.63	0	0	0	2	0	38	30	1	0	0	0	0	6	0
G	47.57	0	0	0	0	0	26	49	0	0	0	0	0	1	0
H	12.75	0	0	18	15	0	2	0	13	0	0	0	20	15	0
I	12.50	15	0	0	0	0	0	0	0	14	0	0	0	0	8
J	0.00	0	0	0	0	0	0	0	0	0	0	98	0	0	0
K	22.94	0	0	0	0	0	0	1	0	0	0	25	0	0	0
L	17.82	0	0	38	17	0	0	0	14	0	0	0	18	7	0
M	20.75	0	0	3	13	0	5	1	19	0	0	0	9	22	0
N	8.42	13	0	0	0	1	0	0	0	15	0	0	0	0	8
O	23.16	20	0	0	0	0	0	0	0	12	0	0	0	0	15
P	9.18	0	0	3	5	0	11	5	8	0	0	0	1	18	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	90	0	0	0
R	51.49	1	0	4	0	0	0	0	0	8	0	0	0	0	8
S	14.14	12	0	0	0	0	0	0	0	13	0	0	0	0	8
T	68.81	17	0	0	0	6	0	0	0	1	0	0	0	0	1
U	34.83	0	0	0	2	0	7	5	17	0	0	0	4	18	0
V	47.92	0	0	0	0	0	0	7	0	0	0	1	0	0	0
W	10.28	0	0	0	0	0	14	45	0	0	0	0	0	0	0
X	0.00	0	0	0	0	0	0	0	0	0	0	65	0	0	0
Y	26.26	0	0	0	0	0	9	53	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	97	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	27.51	110	0	152	79	93	118	236	102	69	0	376	85	106	53

5.2 Variable Selection in Stepwise Linear Discriminant

As explained in Chapter 2, we generated three variables from one-dimensional probability space in relation to the one-graph structure of text files. The expansion from one-dimensional to two-dimensional probability space, that is, from one-graph to digraph structure, has enabled us to generate more variables, *i.e.*, twelve and twenty variables in monoalphabetic and polyalphabetic substitutions, respectively.

During the computation of the coefficients and constants for every group of character, BMDP7M program also selects the variables that have contribution in discriminating the groups. In other words, not all variables in the measurement vectors involve in the computation. Some variables are redundant. Their contribution are not significant because they

Table 5.3 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	31.37	14	0	0	0	2	42	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	4	8	4	0	41	0	0	104
C	47.47	0	1	0	5	0	0	3	0	0	0	0	0	0	99
D	13.21	0	0	0	1	0	0	5	0	0	0	0	0	0	106
E	93.41	0	0	0	0	0	6	0	0	0	0	0	0	0	91
F	33.63	0	0	0	0	0	0	33	0	0	0	3	0	0	11[illegible]
G	47.57	0	1	0	0	0	0	14	0	0	0	12	0	0	103
H	12.75	0	2	0	[illegible]	0	0	16	0	0	0	0	0	0	102
I	12.50	18	0	0	33	15	9	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	2	0	0	0	0	0	100
K	22.94	0	0	0	0	0	0	0	67	10	0	6	0	0	109
L	17.82	0	1	0	4	0	0	2	0	0	0	0	0	0	101
M	20.75	0	11	0	0	0	0	21	0	0	0	0	0	0	106
N	8.42	7	0	0	28	14	9	0	0	0	0	0	0	0	95
O	23.16	22	0	0	6	10	10	0	0	0	0	0	0	0	95
P	9.18	0	9	0	0	0	0	38	0	0	0	0	0	0	98
Q	0.00	0	0	0	0	0	0	0	8	0	0	0	0	0	98
R	51.49	8	0	0	52	20	0	0	0	0	0	0	0	0	101
S	14.14	24	0	0	26	14	2	0	0	0	0	0	0	0	99
T	68.81	7	0	0	1	1	75	0	0	0	0	0	0	0	109
U	34.83	0	5	0	0	0	0	31	0	0	0	0	0	0	89
V	47.92	0	0	0	0	0	0	0	46	10	0	32	0	0	96
W	10.28	0	0	0	0	0	0	2	7	11	0	28	0	0	107
X	0.00	0	0	0	0	0	0	0	29	0	0	1	0	0	95
Y	26.26	0	0	0	0	0	0	3	5	3	0	26	0	0	99
Z	0.00	0	0	0	0	0	0	0	0	1	0	0	0	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
Total	27.51	100	30	0	157	76	153	172	172	39	0	149	0	99	2726

are linearly related to other variables. The values of coefficients and constants of three, twelve, and twenty variables for the groups of character are given in Appendix A.

Stepwise linear discriminant analysis of the training sample presents the following results (see Tables 5.5, 5.6, and 5.7)

(a) all of the three variables generated from one-graph structure made a statistically significant contribution to the computational processes of the discriminant functions in monoalphabetic substitution ;

(b) all of the three variables contributed to the functions in polyalphabetic substitution ;

(c) only eight out of twelve variables generated from digraph structure are significant in monoalphabetic substitution ; this means that we need only to provide values of eight variables in the measurement vectors ; the other four variables are redundant since they are *linearly dependent* upon the eight variables, based on the data provided by the training sample.

In our model of a polyalphabetic substitution cipher there are three cipher alphabets. The first character is enciphered in alphabet one, the second in alphabet two, the third in alphabet three, and so on. "Position one" means the set of all plain or source text characters enciphered in cipher alphabet number one.

(d) in polyalphabetic substitution, the selection processes are different for every position: the first position uses fifteen out of twenty variables, the second position uses fourteen, and the third position uses twelve variables. There are two courses of action if we would like to use the same number of variables for all positions. The first suggestion is to consider the combination "yes, no, no" as a "yes" option. The second suggestion is to consider such a combination as a "no" option. The first one is more conservative and will result in all of twenty variables to be included in the computation of discriminant functions. The second option will use fifteen variables, three-fourth the original number of variables.

5.3 Comparison of Results

One of the objectives of this research is to compare the results of discriminant analysis when applied to both univariate and multivariate situations : how much improvement is there in identifying groups of letter by using the multivariate method compared to univariate method. This research also compares two kinds of discriminant analysis, namely (a) linear discriminant using the stepwise method, and (b) quadratic discriminant analysis.

Tables 5.8, 5.9, 5.10, and 5.11 present a summary of the comparisons in terms of the percentage of correctness of groups classified. In general, it can be said that :

1. Digraph structure provides more variation and improves the total percentage of correctness. It also improves the percentage of correctness for every individual group, in monoalphabetic as well as in polyalphabetic substitutions. This is true in both linear and quadratic discriminant methods.

2. The quadratic discriminant analysis in this research shows some strange results, as can

be seen from the tables. The total percentage of correctness for three variables in all tables are lower than the total percentage of correctness for one variable. This is due to some loss of information during computation because the ranks of covariance matrices are not the same among the groups of characters. The adjustment done to the minimum rank of the matrices according to the minimum number of non-zero eigenvalues of the matrices has caused the trouble (see also the footnote in page 90). This adjustment also responsible for the general inferiority of quadratic discriminant method compared to the linear one. This method is better than the linear method only for analysis with one variable, because there is no loss of information since all matrices are of rank one.

Generally speaking, stepwise linear discriminant analysis is powerful enough in comparison to quadratic discriminant analysis, and can be used appropriately for distinguishing the groups of character.

3. The high percentages of correctness in the analysis of one variable and three variables occur in the groups of characters E and Blank in every substitution and in every method. This means that the variation generated from one-dimensional probability space, or one-graph structure, provides the discriminating power only for these two character groups. By adding more information generated from two-dimensional probability space, or digraph structure, some more characters can be identified and correctly classified. For example, in Table 5.8, linear discriminant analysis of twelve variables shows a high percentage of correctness for character groups D, E, H, L, N, R, S, T, Y and Blank.

5.4 Examples of Recovering Encrypted Texts

This section presents two examples of recovering encrypted source texts by using the linear and quadratic discriminant functions resulting from the computations for one, three, and twelve or twenty variables mentioned in the previous sections. The two source texts are given in Figures 5.1 and 5.2. The source texts are monoalphabetically substituted by using the following rule :

```
ABCDEFGHIJKLMNOPQRSTUVWXYZ@
DEFJKLPQRVWXGHIABCMNOYZ@STU
```

and are polyalphabetically substituted by using the following rules :

```
                ABCDEFGHIJKLMNOPQRSTUVWXYZ@
Position-1  :  GHIJKLMNOPQRSTUVWXYZ@ABCDEF
Position-2  :  JKLMNOPQRSTUVWXYZ@ABCDEFGHI
Position-3  :  UVWXYZ@ABCDEFGHIJKLMNOPQRST
```

The crypto texts resulted from the substitutions are given in Figures 5.3, 5.4, 5.5, and 5.6.

The solution of the two examples by using discriminant functions for one, three, and twelve or twenty variables are given in Appendix C.

The results show that the solutions based on the digraph structures are more readable than the ones based on one-graph structures.

Figures 5.7 to 5.10 show some examples of solution text for monoalphabetic and polyalphabetic substitutions.

Table 5.4: Summary of Tables 5.2 and 5.3 for character groups, monoalphabetic, 1 variable

Group	Training Sample			Validating Sample		
	No.of cases	Correctly classified	Percent correct	No.of cases	Correctly classified	Percent correct
A	98	31	31.63	102	32	31.37
B	96	0	0.00	104	0	0.00
C	101	51	50.50	99	47	47.47
D	94	19	20.21	106	14	13.21
E	109	101	92.66	91	85	93.41
F	87	30	34.48	113	38	33.63
G	97	48	49.48	103	49	47.57
H	98	19	19.39	102	13	12.75
I	88	10	11.36	112	14	12.50
J	100	0	0.00	100	0	0.00
K	91	12	13.19	109	25	22.94
L	99	24	24.24	101	18	17.82
M	94	12	12.77	106	22	20.75
N	105	12	11.43	95	8	8.42
O	105	20	19.05	95	22	23.16
P	102	6	5.88	98	9	9.18
Q	102	0	0.00	98	0	0.00
R	99	62	62.63	101	52	51.49
S	101	12	11.88	99	14	14.14
T	91	48	52.75	109	75	68.81
U	111	34	30.63	89	31	34.83
V	104	32	30.77	96	46	47.92
W	93	4	4.30	107	11	10.28
X	105	0	0.00	95	0	0.00
Y	101	29	28.71	99	26	26.26
Z	102	0	0.00	98	0	0.00
@	101	101	100.00	99	99	100.00
Total	2674	717	26.81	2726	750	27.51

Table 5.5: Variable used in linear discriminant analysis, 3 variables

			Polyalphabetic		
No.	Variable	Monoalphabetic	Position 1	Position 2	Position 3
1.	RELFREQ	yes	yes	yes	yes
2.	INFOCONT	yes	yes	yes	yes
3.	CTOENTRO	yes	yes	yes	yes

Table 5.6: Variable used in linear discriminant analysis, 12 variables

No.	Variable	Monoalphabetic
1.	RELFREQ	yes
2.	ROWCOLSS	yes
3.	WINROWSS	no
4.	WINCOLSS	no
5.	RWCPROSS	yes
6.	CWCPROSS	yes
7.	INFOCONT	yes
8.	CTOENTRO	yes
9.	RCTOJENT	yes
10.	CCTOJENT	yes
11.	RWCENTRO	no
12.	CWCENTRO	no

Table 5.7: Variable used in linear discriminant analysis, 20 variables

No.	Variable	Polyalphabetic			Suggestion-1	Suggestion-2
		Position 1	Position 2	Position 3		
1.	RELFREQ	no	no	yes	yes	no
2.	ROWCOLSS	yes	yes	yes	yes	yes
3.	WINROWS3	yes	no	yes	yes	yes
4.	WINCOLS3	no	no	yes	yes	no
5.	RWCPROS3	yes	no	yes	yes	yes
6.	CWCPROS3	no	yes	yes	yes	yes
7.	INFOCONT	yes	no	no	yes	no
8.	CTOENTRO	yes	yes	yes	yes	yes
9.	RCTOJEN3	yes	yes	no	yes	yes
10.	CCTOJEN3	yes	yes	no	yes	yes
11.	RWCENTR3	no	yes	no	yes	no
12.	CWCENTR3	yes	no	no	yes	no
13.	WINROWSS	yes	yes	no	yes	yes
14.	RWCPROSS	no	yes	yes	yes	yes
15.	RCTOJENT	yes	yes	yes	yes	yes
16.	RWCENTRO	yes	yes	no	yes	yes
17.	WINCOLSS	yes	yes	no	yes	yes
18.	CWCPROSS	yes	yes	yes	yes	yes
19.	CCTOJENT	yes	yes	yes	yes	yes
20.	CWCENTRO	yes	yes	yes	yes	yes
	Total yes	15	14	12	20	15

Table 5.8: Summary of correctness for the validating sample : monoalphabetic substitution

Group	Percent correctness of validating sample					
	Linear Discriminant			Quadratic Discriminant		
	1 var	3 vars	12 vars	1 var	3 vars	12 vars
A	31.37	27.45	61.76	24.00	32.00	50.00
B	0.00	0.00	29.81	0.00	0.00	48.00
C	47.47	47.47	62.63	25.00	23.00	75.00
D	13.21	14.15	84.91	12.00	20.00	82.00
E	93.41	91.21	97.80	96.00	96.00	99.00
F	33.63	41.59	61.95	45.00	51.00	72.00
G	47.57	33.98	57.28	36.00	40.00	47.00
H	12.75	12.75	85.29	16.00	8.00	86.00
I	12.50	12.50	65.18	0.00	1.00	58.00
J	0.00	62.00	34.00	0.00	20.00	20.00
K	22.94	46.79	69.72	51.00	54.00	68.00
L	17.82	15.84	72.25	44.00	55.00	82.00
M	20.75	20.75	42.45	0.00	13.00	51.00
N	8.42	9.47	92.63	21.00	22.00	94.00
O	23.16	22.11	53.68	34.00	30.00	86.00
P	9.18	12.24	50.00	0.00	0.00	51.00
Q	0.00	0.00	67.35	67.00	0.00	70.00
R	51.49	54.46	88.12	71.00	62.00	88.00
S	14.14	15.15	98.99	0.00	0.00	97.00
T	68.81	68.81	88.99	64.00	65.00	95.00
U	34.83	31.46	73.03	53.00	42.00	83.00
V	47.92	47.92	65.63	61.00	50.00	78.00
W	10.28	13.08	48.60	16.00	23.00	61.00
X	0.00	48.42	67.37	39.00	37.00	62.00
Y	26.26	23.23	79.80	24.00	8.00	86.00
Z	0.00	14.29	41.84	0.00	12.00	22.00
@	100.00	100.00	100.00	100.00	100.00	100.00
Total	27.51	32.65	68.09	33.30	32.00	70.78

Table 5.9: Summary of correctness for the validating sample : polyalphabetic substitution, position 1

Group	Percent correctness of validating sample					
	Linear Discriminant			Quadratic Discriminant		
	1 var	3 vars	20 vars	1 var	3 vars	20 vars
A	20.59	22.55	65.69	14.00	15.00	58.00
B	0.00	0.00	40.38	0.00	0.00	27.00
C	44.44	45.45	52.53	25.00	26.00	44.00
D	0.00	0.94	77.36	0.00	0.00	72.00
E	85.71	82.42	96.70	87.00	88.00	96.00
F	6.19	23.01	34.51	11.00	18.00	41.00
G	40.78	29.13	44.66	33.00	37.00	39.00
H	15.69	9.80	80.39	0.00	0.00	85.00
I	18.75	18.75	72.32	0.00	0.00	85.00
J	0.00	0.00	38.00	66.00	0.00	7.00
K	0.00	39.45	39.45	24.00	37.00	29.00
L	30.69	29 70	71.29	42.00	51.00	80.00
M	0.00	3.77	27.36	11.00	20.00	30.00
N	1.05	3.16	87 37	7.00	12.00	92.00
O	17.89	16.84	73.68	32.00	24.00	75.00
P	3.06	6.12	38.78	0.00	0.00	13.00
Q	0.00	0.00	26.53	0.00	0.00	19.00
R	33.66	33.66	68.32	40.00	38.00	72.00
S	14.14	14.14	93.94	13.00	14.00	90.00
T	56.88	54.13	77.06	43.00	50.00	84.00
U	31.46	17.98	71.91	46.00	32.00	68.00
V	15.63	35.42	44.79	41.00	22.00	45.00
W	11.21	16.82	27 10	15.00	16.00	42.00
X	0.00	28.42	45.26	22.00	3.00	35.00
Y	21.21	16.16	70.71	27.00	15.00	66.00
Z	0.00	48.98	48.98	0.00	44.00	71.00
@	98.99	98.99	98.99	100.00	99.00	98.00
Total	20.73	25.57	59.39	25.89	24.48	57.89

Table 5.10: Summary of correctness for the validating sample : polyalphabetic substitution, position 2

Group	Percent correctness of validating sample					
	Linear Discriminant			Quadratic Discriminant		
	1 var	3 vars	20 vars	1 var	3 vars	20 vars
A	27.45	27.45	55.88	24.00	24.00	62.00
B	0.00	0.00	28.85	0.00	0.00	14.00
C	48.48	45.45	57.58	23.00	23.00	52.00
D	26.42	26.42	69.81	46.00	24.00	81.00
E	85.71	85.71	95.60	91.00	93.00	90.00
F	25.66	24.78	37.17	42.00	34.00	57.00
G	24.27	27.18	48.54	17.00	14.00	25.00
H	24.51	20.59	84.31	5.00	16.00	86.00
I	8.93	15.18	74.11	26.00	21.00	75.00
J	0.00	0.00	31.00	37.00	0.00	7.00
K	0.00	52.29	40.37	47.00	35.00	34.00
L	8.91	7.92	71.29	13.00	36.00	75.00
M	12.26	13.21	26.42	0.00	0.00	10.00
N	10.53	8.42	95.79	0.00	2.00	89.00
O	25.26	27.37	67.37	26.00	19.00	58.00
P	0.00	0.00	43.88	0.00	0.00	29.00
Q	0.00	0.00	23.47	0.00	0.00	32.00
R	48.51	49.50	72.28	59.00	53.00	68.00
S	8.08	8.08	87.88	0.00	0.00	86.00
T	49.54	51.38	77.98	47.00	50.00	83.00
U	23.60	24.72	75.28	45.00	38.00	63.00
V	16.67	45.83	60.42	46.00	42.00	58.00
W	0.00	0.00	43.93	22.00	24.00	42.00
X	0.00	43.16	42.11	29.00	1.00	33.00
Y	29.29	33.33	80.81	9.00	0.00	70.00
Z	0.00	44.90	44.90	44.00	43.00	66.00
@	100.00	98.99	100.00	99.00	99.00	99.00
Total	22.12	28.69	60.23	29.52	25.59	57.56

Table 5.11: Summary of correctness for the validating sample : polyalphabetic substitution, position 3

Group	Percent correctness of validating sample					
	Linear Discriminant			Quadratic Discriminant		
	1 var	3 vars	20 vars	1 var	3 vars	20 vars
A	24.51	24.51	53.92	16.00	30.00	79.00
B	0.00	0.00	16.35	12.00	6.00	27.00
C	38.38	37.37	47.47	9.00	5.00	45.00
D	24.53	23.58	77.36	57.00	48.00	68.00
E	84.62	84.62	94.51	83.00	88.00	94.00
F	19.47	27.43	38.94	28.00	25.00	67.00
G	46.60	29.13	44.66	33.00	36.00	11.00
H	18.63	16.67	78.43	1.00	10.00	78.00
I	2.68	8.04	70.54	18.00	14.00	59.00
J	0.00	0.00	39.00	77.00	0.00	1.00
K	0.00	39.45	45.87	55.00	30.00	41.00
L	8.91	7.92	59.41	30.00	45.00	66.00
M	19.81	17.92	23.58	0.00	10.00	21.00
N	2.11	1.05	81.05	2.00	0.00	87.00
O	23.16	24.21	73.68	20.00	23.00	70.00
P	0.00	0.00	31.63	0.00	0.00	33.00
Q	0.00	0.00	39.80	0.00	0.00	40.00
R	44.55	44.55	72.28	53.00	48.00	72.00
S	19.19	17.17	91 92	0.00	0.00	88.00
T	55.96	59.63	80.73	57.00	54.00	81.00
U	29.21	25.84	64.04	37.00	28.00	70.00
V	19.79	48.96	54.17	45.00	44.00	55.00
W	7.48	17.76	39.25	0.00	1.00	16.00
X	0.00	48.42	44.21	8.00	13.00	30.00
Y	0.00	10.10	76.77	24.00	0.00	80.00
Z	0.00	48.98	44.90	0.00	41.00	67.00
@	100.00	100.00	100.00	100.00	100.00	100.00
Total	21.61	28.03	58.36	28.33	25.89	57.26

Figure 5.1: Source text, example 1

LETTERS DEAR EDITOR THANK YOU FOR THE ARTICLE IN YOUR MAY IS
SUE IS COURSE TARGETED AT MANAGERS WHICH DESCRIBES THE COMPU
TER BASED INFORMATION SYSTEMS PROGRAM AT THE UNIVERSITY OF V
ICTORIA THERE IS HOWEVER AN ERROR IN THE THIRD PARAGRAPH OF
THAT ARTICLE CURRENTLY THERE ARE APPROXIMATELY ACTIVE STUDEN
TS IN THE COMPUTER BASED INFORMATION SYSTEMS PROGRAM AS OPPO
SED TO THE STUDENTS MENTIONED IN THE ARTICLE THE TOTAL NUMBE
R OF STUDENTS WHO HAVE REGISTERED FOR THE PROGRAM SINCE ITS
INCEPTION IN IS APPROXIMATELY OF THOSE STUDENTS HAVE COMPLET
ED THE PROGRAM THE COMPUTING TOOLS FOR MANAGEMENT COURSE DIS
CUSSED IN THE ARTICLE HAD AN ENROLLMENT OF STUDENTS AS MENTI
ONED IT IS SOME OF NINE COURSES WHICH COMPRISE THE CBIS PROG
RAM SINCERELY JEANETTE MUZIO PROGRAM COORDINATOR COMPUTER BA
SED INFORMATION SYSTEMS UNIVERSITY OF VICTORIA VICTORIA BC L
ETTERS DEAR EDITOR I WAS PLEASED TO NOTE SYDNEY DEVELOPMENT
CORPORATIONS PRESS RELEASE ANNOUNCING OUR AGREEMENT WITH UNI
SYS CORPORATION IN YOUR MAY ISSUE HOWEVER YOU HAD ME QUICKLY
SHAKING MY HEAD WHEN YOU STATED THAT SYDNEYS MESSENGER AND
RFC GATEWAY SOFTWARE WOULD BE INCORPORATED INTO UNISYSS ELEC
TRONIC MASSAGING PRODUCTS RATHER THAN MESSAGING PRODUCTS IS
THIS A NEW PRODUCT LINE WE ARE UNAWARE OF WHILE HUMOROUS SUC
H AN ERROR IS A SOURCE OF CONFUSION YOURS SINCERELY BARBARA
MERLO MARKETING MANAGER SYDNEY DEVELOPMENT CORP VANCOUVER BC
LETTERS DEAR EDITOR I WANT TO CLARIFY NCR COMTENS INVOLVEME
NT IN THE MOUNTAIN BELL ISDN TRIAL AS DESCRIBED BY A MARCH C
OMPUTING CANADA ARTICLE ISDN EDGING CLOSER TO REALITY NCR CO
MTEN PROVIDED COMMUNICATIONS PROCESSORS AND NCR CORPORATIONS
PERSONAL COMPUTER DIVISION PROVIDED PERSONAL COMPUTERS FOR
MOUNTAIN BELLS ISDN TRIAL NCR COMTEN AND THE NCR PC DIVISION
DEVELOPED TWO ISDN TERMINAL ADAPTERS FOR THE TRIAL ONE INTE
GRATED INTO AN NCR PC THE SECOND A STANDALONE ADAPTER THIS E
QUIPMENT WAS USED TO MAKE THE FIRST CUSTOMER APPLICATION CAL
L OVER THE MOUNTAIN BELL ISDN TRIAL NETWORK DURING THE NEWS
CONFERENCE THAT SIGNALED THE BEGINNING OF THE TRIAL MICHELE
WOLFF PUBLIC RELATIONS REPRESENTATIVE CORPORATE COMMUNICATIO
NS NCR COMTEN ST PAUL MINN

Figure 5.2 Source text, example 2

EDITORIAL STANDARDS UNFOLD AS THEY SHOULD INVARIABLY STANDAR
DS MOVEMENTS SEEM TO GROW BY LEAPS AND BOUNDS IT IS ALL TOO
PREDICTABLE IN THE EARLY STAGES A FEW VISIONARIES SPREAD THE
 MESSAGE WITH AN EVANGELICAL ZEAL USERS ARE IMPRESSED AND SE
EK WAYS TO APPLY THE NEW CONCEPTS TO THEIR APPLICATIONS BUT
THEN DISILLUSIONMENT SETS IN SYSTEM CONVERSION AND CUSTOMIZA
TION ALWAYS PROVES TO BE A BIGGER JOB THAN ANYONE INCLUDING
MIS COULD HAVE PREDICTED COSTS SOAR VENDORS PRODUCTS NEVER S
EEM TO QUITE LIVE UP TO THE PROMISES USERS REVERT TO THE TRI
ED AND TRUE AND DECIDE TO LEAVE THE RISKS TO THE PIONEERS TH
IS SEEMS TO BE THE CASE WITH AT LEAST THREE OF THE HIGHER PR
OFILE STANDARDSMAKING EFFORTS ELECTRONIC DATA INTERCHANGE ED
I SEE STORY P MANUFACTURING AUTOMATION PROTOCOLTECHNICAL OFF
ICE PROTOCOL MAPTOP SEE STORY P AND OPEN SYSTEMS INTERCONNEC
TION OSI SEE STORY P IN EACH CASE THE MOVEMENT NEVER SPREAD
QUITE AS QUICKLY AS FIRST ANTICIPATED EDI WHICH PROPOSES TO
MOVE BUSINESS DOCUMENTS SUCH AS INVOICES AND PURCHASE ORDERS
 ELECTRONICALLY BETWEEN COMPANIES PROVED SLOWER AND MORE EXP
ENSIVE THAN EXPECTED EARLY ENTHUSIASM FOR MAP STANDARDIZED C
OMMUNICATIONS FOR DEVICES ON THE FACTORY FLOOR HAS GIVEN WAY
 TO SKEPTICISM MOSTLY BECAUSE OF THE SLOW DEVELOPMENT OF MAP
 STANDARDS AND PRODUCTS AND OSI A GLOBAL EFFORT TO ACHIEVE I
NTEROPERABILITY BETWEEN A WIDE RANGE OF COMPUTER SYSTEMS FAC
ES NATIONAL CONFLICTS BUT IS THAT ANY REASON FOR USERS TO FR
ET AT THE SLOW PACE ACCEPTANCE OF STANDARDS NOT AT ALL STAND
ARDS REMAIN A NECESSARY PART IN THE EVOLUTION OF SYSTEMS REG
ARDLESS OF HOW SLOWLY THE FIRST STEPS ARE TAKEN THE EXPERTS
HAVE REVISED THEIR FORECASTS AND WHILE ACKNOWLEDGING THAT AC
CEPTANCE IS COMING SLOWER THAN FIRST ANTICIPATED IT IS WORTH
 NOTING THAT THEIR OPTIMISM REMAINS SURVEYS INDICATE TOO THA
T USERS INTEREST IN STANDARDS IS WELL ESTABLISHED ITS JUST T
HAT MOST OF THEM HAVE DECIDED TO WAIT FOR EXAMPLE WHILE MUCH
 OF THE WHOLEHEARTED ENTHUSIASM IS GONE AND SOME QUESTIONED
THE VIABILITY OF MAP PRODUCTS THE UNDERLYING NEED FOR AN IND
USTRY STANDARD ON THE FACTORY FLOOR REMAINS SOME OF THE PROB
LEM WILL BE SOLVED IN TIME THROUGH EDUCATION OSI WATCHERS IN
 THE FEDERAL GOVERNMENT NOTE DIFFERENT LEVELS OF KNOWLEDGE A
MONG DIFFERENT USERS THEIR ADVICE IS THAT FOR NOW USERS STAY
 UP TO DATE WITH OSI DEVELOPMENTS LITERATURE AND SO ON OR TH
EY CAN FIND THEMSELVES QUICKLY FALLING BEHIND FOR ALL THE EX
PECTED SHORT TERM PROBLEMS AND COSTS ASSOCIATED WITH THE EAR
LY IMPLEMENTATION OF STANDARDS THE LONG TERM GAINS STILL MAK
E IT ALL WORTHWHILE REDUCED RESEARCH AND DEVELOPMENT COSTS F
EWER GATEWAYS AND THEIR ASSOCIATED COSTS AND REDUCED PERFORM
ANCE AND CONFORMANCE TESTING

Figure 5.3: Crypto text, example 1, monoalphabetic substitution

```
XKNNKCMUJKDCUKJRNICUNQDHWUSIOULICUNQKUDCNRFXKURHUSIOCUGDSURM
MOKURMUFIOCMKUNDCPKNKJUDNUGJHDPKCMUZQRFQUJKMFCREKMUNQKUFIGAO
NKCUEDMKJURHLICGDNRIHUMSMNKGMUACIPCDGUDNUNQKUOHRYKCMRNSUILUY
RFNICRDUNQKCKURMUQIZKYKCUDHUKCCICURHUNQKUNQRCJUADCDPCDAQUILU
NQDNUDCNRFXKUFOCCKHNXSUNQKCKUDCKUDAACI RGDNKXSUDFNRYKUMNOJKH
NMURHUNQKUFIGAONKCUEDMKJURHLICGDNRIHUMSMNKGMUACIPCDGUDMUIAAI
MKJUNIUNQKUMNOJKHNMUGKHNRIHKJURHUNQKUDCNRFXKUNQKUNINDXUHOGEK
CUILUMNOJKHNMUZQIUQDYKUCKPRMNKCKJULICUNQKUACIPCDGUMRHFKURNMJ
RHFKANRIHURHURMUDAACI RGDNKXSUILUNQIMKUMNOJKHNMUQDYKUFIGAXKN
KJUNQKUACIPCDGUNQKUFIGAONRHPUNIIXMULICUGDHDPKGKHNUFIOCMKUJRM
FOMMKJURHUNQKUDCNRFXKUQDJUDHUKHCIXXGKHNUILUMNOJKHNMUDMUGKHNR
IHKJURNURMUMIGKUILUHRHKUFIOCMKMUZQRFQUFIGACRMKUNQKUFERMUACIF
CDGUMRHFKCKXSUVKDHKNNKUGOTRIUACIPCDGUFIICJRHDNICUFIGAONKCUED
MKJURHLICGDNRIHUMSMNKGMUOHRYKCMRNSUILUYRFNICRDUYRFNICRDUEFUX
KNNKCMUJKDCUKJRNICURUZDMUAXKDMKJUNIUHINKUMSJHKSUJKYKXIAGKHNU
FICAICDNRIHMUACKMMUCKXKDMKUDHHIOHFRHPUIOCUDPCKKGKHNUZRHQUOHR
MSMUFICAICDNRIHUKHUSIOCUGDSURMMOKUQIZKYKCUSIOUQDJUGKUBORFWXS
UMQDWRHPUGSUQKDJUZQKHUSIOUMNDNKJUNQDNUMSJHKSMUGKMMKHPKCUDHJU
CLFUPDNKZDSUMILNZDCKUZIOXJUEKURHFICAICDNKJURHNIUOHRMSMMUXXKF
NCIHRFUGDMMDPRHPUACIJOFNMUCDNQKCUNQDHUGKMMDPRHPUACIJOFNMURMU
NQRMUDUHKZUACIJOFNUXRHKUZKUDCKUOHDZDCKUILUZQRXKUQOGICIOMUMOF
QUDHUKCCICURMUDUMIOCFKUILUFIHLOMRIHUSIOCMUMRHFKCKXSUEDCEDCDU
GKCXIUGDCWKNRHPUGDHDPKCUMSJHKSUJKYKXIAGKHNUFICAUYDHFIOYKCUEF
UXKNNKCMUJKDCUKJRNICURUZDHNUNIUFXDCRLSUHFCUFIGNKHMURHYIXYKGK
HNURHUNQKUGIOHNDRHUEKXXURMJHUNCRDXUDMUJKMFCREKJUESUDUGDCFQUF
IGAONRHPUFDHDJDUDCNRFXKURMJHUKJPRHPUFXIMKCUNIUCKDXRNSUHFCUFI
GNKHUACIYRJKJUFIGGOHRFDNRIHMUACIFKMMICMUDHJUHFCUFICAICDNRIHM
UAKCMIHDXUFIGAONKCUJRYRMRIHUACIYRJKJUAKCMIHDXUFIGAONKCMULICU
GIOHNDRHUEKXXMURMJHUNCRDXUHFCUFIGNKHUDHJUNQKUHFCUAFUJRYRMRIH
UJKYKXIAKJUNZIURMJHUNKCGRHDXUDJDANKCMULICUNQKUNCRDXUIHKURHNK
PCDNKJURHNIUDHUHFCUAFUNQKUMKFIHJUDUMNDHJDXIHKUDJDANKCUNQRMUK
HORAGKHNUZDMUOMKJUNIUGDWKUNQKULRCMNUFOMNIGKCUDAAXRFDNRIHUFDX
XUIYKCUNQKUGIOHNDRHUEKXXURMJHUNCRDXUHKNZICWUJOCRHPUNQKUHKZMU
FIHLKCKHFKUNQDNUMRPHDXKJUNQKUEKPRHHRHPUILUNQKUNCRDXUGRFQKXKU
ZIXLLUAOEXRFUCKXDNRIHMUCKACKMKHNDNRYKUFICAICDNKUFIGGOHRFDNRI
HMUHFCUFIGNKHUMNUADOXUGRHHU
```

Figure 5.1: Crypto text, example 1, polyalphabetic substitution

```
123123123123123123123123123123123123123123123123123123123123

RNMZNKYIXKJKFNXOBHXIMNJGQIRUCTLXKFBAKIUXBBIUYFRGFGH  TSJRFRL
YCYFRLFLH  LKIMG  KBYJIUZIFGWUMNKYIPNRWNIXKAWXRVKATZQYFLHSYN
ZNKFKUYNXFRGLXKSJMOXGFARYBYSATV HM USIUZIMNNT WBANKYRMDIHLIO
OLMU BGIMNNKKIBYIAUEYANKFJGFNKXXKFRGFBAKIMNRKJIIG UM UVQTUOT
ZQUZIUXBBIUYFLNX YTBEDIMNNKKIUXNTGYIXXQOVUZNEDIUIBBANTYBNJNG
ZATOWTZQYFLHSYNZNKFKUYNXFRGLXKSJMOXGFARYBYSATV HM USIUYIHVYH
YNXFBHFBAKILZCXKWMYIFKWMOXGKMTOWTZQYFJKZRWRNTZQYFBHZJEFWNSKY
XIHLILZCXKWMYIPNXTNJOKIKKPBYBYXNXFOHXIMNNTV HM USILOWWKIBZAT
OWWKYMOXGFRGFRLFJIV HCRFGBYRGTUOTZQHYNTYBNJNGZATNJOKIWUVIRNM
KMTZQYFYKUPKGVTZQYFLHSYNZRGMIMUXEYIZU TSJGGPYSNGZIWUCKYNTJRL
ICLYNXFRGFBAKIUXBBIUYFQUJIUTIYT HRUFKWMFXZFAM MYTBLFJLFVYTBB
UWYJIBZIBYILUVYFXZFWBTNTIXNXAYYIPNRWNIWUVIXRLKIMNNTIKBYIIXX
XJFFABTLYXNEDICKJGKBMKIF HBUIIXX XJFFLHU XOWUZXKFLHSYNZNKFKU
YNXFRGLXKSJMOXGFARYBYSAT WBANKYRMDIHLIOOLMU BGIOOLMU BGIVIIE
KBMK LFMYG TKMBZXKFRTBJLFYEKJLVMTZXTTXMKILDMGKGTJNOKUHVVYTBT
IXKVXKGBBUWLFYKKALF YRNUYNTGWGUCGIRGMIH  TGPKKNFKWMFEBZQT WB
YGLFLHXYHXJMOXGFRGFGH  TSJRFRLYCYFQHBNOK TDXNFQUJIFKIJ RWQUR
FAAGTBTPTSGTNNUJIPNNGFGH ILZJMKMTZQUZILDMGKGLFVYYAYTPYXIUTMT
XOWFPUZNPGGTYXZZEUXNTBXNRMTHNTOWWU IU UZNXFRGZXT WBYGLYIYRNW
Z HTRWFVUYAUMRGMIIXXX LMYIKGBAK TZQUTIFKALGPBTPTV HJCWZATOAT
ZQBYIUFWYBIIXXX LMFUBTNTBNTG YFCGGEUXNTUOTBQBRNTNCFU H ATYCW
NIUTIYX HXIBYIUFAH  WKIHLIWUWZ ABUWTDXNXATYRGINKKURFKUXKUXJT
SNKRXTSJKQNMOW FVUTJ K TYGXTNRFMYANEUYFKWMFLHXYTAJGIXNANKFKW
FUYZBYXATJNUXIYJRMU TOIPGWMFBHFLEG BLGTTLKFLHSBYTATOWOUUOKVY
TBTOWTZQYFVH WMGRGFKYRUTOAXTIMXRURIUYIXKAWXRVKMTHGTGIFG WNIW
UVI BBTPTIJGGMUFJKZRWRNTOAXTIYJPBTPTIUHYNKFBHF YGUBZGTTLKFLH
SBYTIIXXOOMYJIWUVF WBIJMOXGYIIXXWKALU LFJGJIGI TIXKVXKGBBUWL
FYYXAHTJEFLHSYNZNKFMBARLOXGFYKUDBJNXFYYXAHTJEFLHSYNZNKYIZU T
SXNTBUOWTHNERATOAXTIMXRURIGI TIXFZNGFJGJIMNNTTLKFYWFMBARLOXG
IMYANEUYYJIMBXTOAXTIMK FOWURIUJJIZNKYIZU TZQYFBKOJEFXGKIBTBY
M UZNXFRGZXTGWTTLKFYWFBAKILKLHTMTGILZJGJJEUWYFJXGYMK TZQBYIY
WCBVVYTBTBJLFCLKMTZXTSJDKIMNNTLRKYBTICLZXFK TGYIRRWGBBUWTIJE
RIHANKFBAKIFUCGZJBTIVKUEFRLJWTZ BGUTTNMBXKQIX  BTPTZQYFWYBAT
IXGLNKKWWKIMNJMFABMWURNXFBAKIVKPBTWBTPTUOTZQYFBKOJEFVBIQYRNT
BXELOTVCVRRWF YRJMOXGYIKKYKKAYTBUZROKIWU IU UZNTIXFSCGOLUZRH
TATTLKFLHSBYTILZIIGCEFVBTWT
```

Figure 5.5: Crypto text, example 2, monoalphabetic substitution

```
KJRNICRD~UMNDHJDCJMUOHLIXJUDMUNQKSUMQIOXJURHYDCRDEXSUMNDHJDC
JMUGIYKGKHNMUMKKGUNIUPCIZUESUXKDAMUDHJUEIOHJMURNURMUDXXUNIIU
ACKJRFNDEX: JRHUNQKUKDCXSUMNDPKMUDULKZUYRMRIHDCRKMUMACKDJUNQK
UGKMMDPKUZRNQUDHUKYDHPKXRFDXUTKDXUOMKCMUDCKURGACKMMKJUDHJUMK
KWUZDSMUNIUDAAXSUNQKUHKZUFIHFKANMUNIUNQKRCUDAAXRFDNRIHMUEONU
NQKHUJRMRXXOMRIHGKHNUMKNMURHUMSMNKGUFIHYKCMRIHUDHJUFOMNIGRTD
NRIHUDXZDSMUACIYKMUNIUEKUDUERPPKCUVIEUNQDHUDHSIHKURHFXOJRHPU
GRMUFIOXJUQDYKUACKJRFNKJUFIMNMUMIDCUYKHJICMUACIJOFNMUHKYKCUM
KKGUNIUBORNKUXRYKUOAUNIUNQKUACIGRMKMUOMKCMUCKYKCNUNIUNQKUNCR
KJUDHJUNCOKUDHJUJKFRJKUNIUXKDYKUNQKUCRMWMUNIUNQKUARIHKKCMUNQ
RMUMKKGMUNIUEKUNQKUFDMKUZRNQUDNUXKDMNUNQCKKUILUNQKUQRPQKCUAC
ILRXKUMNDHJDCJMGDWRHPUKLLICNMUKXKFNCIHRFUJDNDURHNKCFQDHPKUKJ
RUMKKUMNICSUAUGDHOLDFNOCRHPUDONIGDNRIHUAC!NIFIXNKFQHRFDXUILL
RFKUACINIFIXUGDANIAUMKKUMNICSUAUDHJUIAKHUMSMNKGMURHNKCFIHHKF
NRIHUIMRUMKKUMNICSUAURHUKDFQUFDMKUNQKUGIYKGKHNUHKYKCUMACKDJU
BORNKUDMUBORFWXSUDMULRCMN''DHNRFRADNKJUKJRUZQRFQUACIAIMKMUNIU
GIYKUEOMRHKMMUJIFOGKHNMUMOFQUDMURHYIRFKMUDHJUAOCFQDMKUICJKCM
UKXKFNCIHRFDXXSUEKNZKKHUFIGADHRKMUACIYKJUMXIZKCUDHJUGICKUK A
KHMRYKUNQDHUK AKFNKJUKDCXSUKHNQOMRDMGULICU'GDAUMNDHJDCJRTKJUF
IGGOHRFDNRIHMULICUJKYRFKMUIHUNQKULDFNICSULXIICUQDMUPRYKHUZDS
UNIUMWKANRFRMGUGIMNXSUEKFDOMKUILUNQKUMXIZUJKYKXIAGKHNUILUGDA
UMNDHJDCJMUDHJUACIJOFNMUDHJUIMRUDUPXIEDXUKLLICNUNIUDFQRKYKUR
HNKCIAKCDERXRNSUEKNZKKHUDUZRJKUCDHPKUILUFIGAONKCUMSMNKGMULDF
KMUHDNRIHDXUFIHLXRFNMUEONURMUNQDNUDHSUCK9MIHULICUOMKCMUNIULC
KNUDNUNQKUMXIZUADFKUDFFKANDFFKUILUMNDHJDCJMUHINUDNUDXXUMNDHJ
DCJMUCKGDRHUDUHKFKMMDCSUADCNURHUNQKUKYIXONRIHUILUMSMNKGMUCKP
DCJXKMMUILUQIZUMXIZXSUNQKULRCMNUMNKAMUDCKUNDWKHUNQKUK AKCNMU
QDYKUCKYRMKJUNQKRCULICKFDMNMUDHJUZQRXKUDFWHIZXKJPRHPUNQDNUDF
FKANDHFKURMUFIGRHPUMXIZKCUNQDHULRCMNUDHNRFRADNKJURNURMUZICNQ
UHINRHPUNQDNUNQKRCUIANRGRMGUCKGDRHMUMOCYKSMURHJRFDNKUNIIUNQD
NUOMKCMURHNKCKMNURHUMNDHJDCJMURMUZKXXUKMNDEXRMQKJURNMUVOMNUN
QDNUGIMNUILUNQKGUQDYKUJKFRJKJUNIUZDRNULICUK DGAXKUZQRXKUGOFQ
UILUNQKUZQIXKQKDCNKJUKHNQOMRDMGURMUPIHKUDHJUMIGKUBOKMNRIHKJU
NQKUYRDERXRNSUILUGDAUACIJOFNMUNQKUOHJKCXSRHPUHKKJULICUDHURHJ
OMNCSUMNDHJDCJUIHUNQKULDFNICSULXIICUCKGDRHMUMIGKUILUNQKUACIE
XKGUZRXXUEKUMIXYKJURHUNRGKUNQCIOPQUKJOFDNRIHUIMRUZDNFQKCMURH
UNQKULKJKCDXUPIYKCHGKHNUHINKUJRLLKCKHNUXKYKXMUILUWHIZXKJPKUD
GIHPUJRLLKCKHNUOMKCMUNQKRCUDJYRFKJRMUNQDNULICUHIZUOMKCMUMNDS
UOAUNIUJDNKUZRNQUIMRUJKYKXIAGKHNMUXRNKCDNOCKUDHJUMIUIHUICUNQ
KSUFDHULRHJUNQKGMKXYKMUBORFWXSULDXXRHPUEKQRHJULICUDXXUNQKUK
AKFNKJUMQICNUNKCGUACIEXKGMUDHJUFIMNMUDMMIFRDNKJUZRNQUNQKUKDC
XSURGAXKGKHNDNRIHUILUMNDHJDCJMUNQKUXIHPUNKCGUPDRHMUMNRXXUGDW
KURNUDXXUZICNQZQRXKUCKJOFKJUCKMKDCFQUDHJUJKYKXIAGKHNUFIMNMUL
KZKCUPDNKZDSMUDHJUNQKRCUDMMIFRDNKJUFIMNMUDHJUCKJOFKJUAKCLICG
DHFKUDHJUFIHLICGDHFKUNKMNRHPU
```

Figure 5.6. Crypto text, example 2, polyalphabetic substitution

```
123123123123123123123123123123123123123123123123123123123123123

KMBZXKOJEFAMGWXG XYINTOHRMTGATZQYDILNXNRMTOWOG BGKEDILZJGJJK
JATSXOKVYTBLFAYKVTZXTM HBIVDIEKJIYIUTMTHXNTMLFRMFRLFJERIMUXT
V YJRWZJVRNTOWTZQYFNUXURFAMGPYYIUFOYBIOOABUWUXRYYILV YGMTZQY
FVYYAUMNTBRMNIUTIYAJGMNEOLURISKJEFCLK LFJKKIBSYKKALKMTGWXFAY
KTTBJRYIMUIUVYEDIMNNTTNPFLHTLYVBLFBHFBAKRKFJIVUBIJMOXGYIV BT
ZQYTIXOABRUNYRHTVYTBTYNMYIBTILDAMKVTIXGANKYRHTIUTMTICLZXFOHU
ZRHTIUREUDATV HANLFBHFKYFJTHR MNKFSHHIMNJGFJGDXGKIBTLE MBTPT
SRLFLH UXFQUANTV YJRWZNXFLHYBLFAHG TANGJXKYIIXXX LMYIGKDYXIL
KNFFBHFZNOBYFUBANT YTZXTZQYFYKUVBYNLFCLK LF YANKZIMUIMNNTZ B
KMTGWXFBK NTGWXFMYIRXKIMUIEKJOKIMNNTXRLQATZXTZQYFYBUWYK LFBA
OATYNYSATZXTHNTZQYFLUYNTBRMNIUZIEKJLZIMN YKIHLIMNNTNR NNKFYK
UOBRNTYBUTMUXMLSJDOW FNZLXKZATKUYIBKUWBIIXGBUFRGZNKIQUTPYFNX
OILKNTYBHXGTVIFGWNLJWZCKOW FJNZXFGBBUWTV HZXWUUMKLATRWGUTUOZ
OLYFYXUBHIXEFVUVBHVILKNTYBHXGTVIUTMTUYYTILDAMKVLFRGZNKIXGTNW
ZRHTIHYRTYNYFAMU RFYTOWTKJWNIWGAYFBAKIFUDYSNGZIGKDYXILV YGMT
WCBZNTGATWCBITEDIUYIZO LZIUTBBIRIGBYJIYJRTBQBIQTV HVXLKATZXT
SXOKIV ABTNLYIXULNSNGZATYCWNIUYIBTPHOLYYIUTMTVCKIQUYNTU XK L
FNEKLMXXGOLURURFKYZEYKWTIXFVJGONLFYKUDYJILRXPK TGWXFVHXNTKFI
KWLODYFBAGWTKFIKLMKMTKJKRGTKWMNCLOJLSTZU TSJIFAMGWXG XOHYJIW
UVF WBIJMOXGYIZU TJNOOLYYIHTIMNNTLJWZXKDIZRXHXIAGATMROKWTBJR
FBHFADKYMOLBYVTSXLZURFKYIJNYNTUOTZQYFAEUETJNOKUHVVYTBTUOTSJI
FAMGWXG XYIUTMTV HJCWZATGW XLOPJFPEUKURIYLOHXBTZXTGLAONOKIB
TBYXXIK UHREJBRFKYZEYKWTGIPOMYF UTPYFXZFLHSYNZNKFARYBYSATLJW
KATTJMOXGGUTIXGLUBIBLFKNZIBYIMNJMFJGDIKKJLUWTLXKFCLK LFBHFOK
KBTGBTZQYFAEUETVJWKIUILYVBUTLYFXZFAMGWXG XYIGUBTGBTGUEFAMGWX
G XYIKKVUOWTGIGKLYYAUXGTVJKZIBTIMNNTKDHRCMOXGFXZFARYBYSATXN
G XRNLYIHLIAUETYJHBURFBAKIZO LZILZNIYIUXNTZJDKWTZQYFNQVNKZAT
NJOKIKKDBYNXFBAKRKFOHXNWGAMYIUTMTBQBRNIGLDTXPRNXMRGMIMNJMFJW
INIZJGINTOATIXFOW FAEUEYXIMNJGFOBXAMFJGZRWOYUZNXFRMFRLFEHXBA
FWHZRGMIMNJMFBAKRKFXIZRFOAFF YSJBTATYCKANRYITTMBIJMKIMUXTZQU
ZINYNKYIBTBYXNLZIBTILZJGJJKJATOATBNERIYYBUHUBYQYJIBZATPCLZIM
NJMFVHYBTUOTZQYSIAGDYFMYIRXKMTZXTBJBZIZU TKFUSYEKIPNREKIF LA
FXZFBAKIPNXEKQYG MKMTKWMNCLOJLSIBYI UWYFJGJILUVYFZNKAMOXGKMT
ZQYFDBGKBRRMDIHLIFGYTV HJCWZATZQYFCGJNKRGBTPTTNYJIZU TGWTOWX
 AMXGTYBUTMUXMTUWTZQYFOUIBHXGTLUHU TXNFGRGYILUVYFXZFBAKIIXXV
RNFFEBRUTHNTYXEANXFRGFBBSNTZQKUC NIYJCWGBBUWTUABFEUZLAK LFRG
FBAKIZKMYXJEFPHANKTVYTBTTXMKIXOOZK YTBTRNOKULFXZFTGUEEKM KIU
SXGMIXOOZK YTBT AYXATZQYO TGMOOLYFRLFBAGBTLXKFWHBINYNKYILZJR
FCIFBHFMUZNTBRMNIHYRTJNOKUHVVYTBLFUBZNKGBNXNTGWXFAHFXGFXKFBA
KGTIJGFOBTMTZQYSAYRDYYIJ RWQURFOURUBTPTHNAOWXFOHXIURUTZQYFNQ
VNWZNXFAAU MFBYXVTV HHUYSATGWXFLHYBLFJLYXWOJMKMTBRMNIMNNTKJK
RGTOVIRNFKWMGBBUWTJOTYBUTMUXMLFBAKIEUW FBYXVTMJBTATYBBRUTSJD
KIBZIURUTBXKZQPNREKIKKMNINXF YYNUXLAFJGJIXKDYRXISNGZIWUAMYIZ
KEYXI GBYBJRYIUTMTZQYO TGALULBGBYJIWUAMYIUTMTXNX LYJIIK ZU F
GWWKIUTMTIXGLXKSJGINTZNLZRGMI
```

Figure 5.7: Cryptanalysis of monoalphabetic substitution using linear discriminant, 12 variables, example 2

```
EDITORIIL STINDIRDS UNFOLD IS THEY SHOULD INAIRIIWLY STINDIR
DS MOAEMENTS SEEM TO GROW WY LEIPS IND WOUNDS IT IS ILL TOO
PREDICTIWLE IN THE EIRLY STIGES I FEW AISIONIRIES SPREID THE
 MESSIGE WITH IN EAINGELICIL JEIL USERS IRE IMPRESSED IND SE
EK WIYS TO IPPLY THE NEW CONCEPTS TO THEIR IPPLICITIONS WUT
THEN DISILLUSIONMENT SETS IN SYSTEM CONAERSION IND CUSTOMIJI
TION ILWIYS PROAES TO WE I WIGGER JOW THIN INYONE INCLUDING
MIS COULD HIAE PREDICTED COSTS SOTR AENDORS PRODUCTS NEAER S
EEM TO QUITE LIAE UP TO THE PROMISES USERS REAERT TO THE TRI
ED IND TRUE IND DECIDE TO LEIAE THE RISKS TO THE PIONEERS TH
IS SEEMS TO WE THE CISE WITH IT LEIST THREE OF THE HIGHER PR
OFILE STINDIRDSMIKING EFFORTS ELECTRONIC DITI INTERCHINGE ED
I SEE STORY P MINUFICTURING IUTOMITION PROTOCOLTECHNICIL OFF
ICE PROTOCOL MIPTOP SEE STORY P IND OPEN SYSTEMS INTERCONNEC
TION OSI SEE STORY P IN EICH CISE THE MOAEMENT NEAER SPREID
QUITE IS QUICKLY IS FIRST INTICIPITED EDI WHICH PROPOSES TO
MOAE WUSINESS DOCUMENTS SUCH IS INAOICES IND PURCHISE ORDERS
 ELECTRONICILLY WETWEEN COMPINIES PROAED SLOWER IND MORE EXP
ENSIAE THIN EXPECTED EIRLY ENTHUSIISM FOR MIP STINDIRDIJED C
OMMUNICITIONS FOR DEAICES ON THE FICTORY FLOOR HIS GIAEN WIY
 TO SKEPTICISM MOSTLY WECIUSE OF THE SLOW DEAELOPMENT OF MIP
 STINDIRDS IND PRODUCTS IND OSI I GLOWIL EFFORT TO ICHIEAE I
NTEROPERIWILITY WETWEEN I WIDE RINGE OF COMPUTER SYSTEMS FIC
ES NITIONIL CONFLICTS WUT IS THIT INY REISON FOR USERS TO FR
ET IT THE SLOW PICE ICCEPTINCE OF STINDIRDS NOT IT ILL STIND
IRDS REMIIN I NECESSIRY PIRT IN THE EAOLUTION OF SYSTEMS REG
IRDLESS OF HOW SLOWLY THE FIRST STEPS IRE TIKEN THE EXPERTS
HIAE REAISED THEIR FORECISTS IND WHILE ICKNOWLEDGING THIT IC
CEPTINCE IS COMING SLOWER THIN FIRST INTICIPITED IT IS WORTH
 NOTING THIT THEIR OPTIMISM REMIINS SURAEYS INDICITE TOO THI
T USERS INTEREST IN STINDIRDS IS WELL ESTIWLISHED ITS JUST T
HIT MOST OF THEM HIAE DECIDED TO WIIT FOR EXIMPLE WHILE MUCH
 OF THE WHOLEHEIRTED ENTHUSIISM IS GONE IND SOME QUESTIONED
THE IIIWILITY OF MIP PRODUCTS THE UNDERLYING NEED FOR IN IND
USTRY STINDIRD ON THE FICTORY FLOOR REMIINS SOME OF THE PROW
LEM WILL WE SOLAED IN TIME THROUGH EDUCITION OSI WITCHERS IN
 THE FEDERIL GOAERNMENT NOTE DIFFERENT LEAELS OF KNOWLEDGE I
MONG DIFFERENT USERS THEIR IDAICE IS THIT FOR NOW USERS STIY
 UP TO DITE WITH OSI DEAELOPMENTS LITERITURE IND SO ON OR TH
EY CIN FIND THEMSELAES QUICKLY FILLING WEHIND FOR ILL THE EX
PECTED SHORT TERM PROWLEMS IND COSTS ISSOCIITED WITH THE EIR
LY IMPLEMENTITION OF STINDIRDS THE LONG TERM GIINS STILL MIK
E IT ILL WORTHWHILE REDUCED RESEIRCH IND DEAELOPMENT COSTS F
EWER GITEWIYS IND THEIR ISSOCIITED COSTS IND REDUCED PERFORM
INCE IND CONFORMINCE TESTING
```

Figure 5.8: Cryptanalysis of monoalphabetic substitution by using quadratic discriminant, 12 variables, example 2

```
EDITORIAL STANDARDS UNFOLD AS THEY SHOULD INYARIAWLY STANDAR
DS MOYEMENTS SEEM TO GROW WY LEAPS AND WOUNDS IT IS ALL TOO
PREDICTAWLE IN THE EARLY STAGES A FEW YISIONARIES SPREAD THE
 MESSAGE WITH AN EYANGELICAL ZEAL USERS ARE IMPRESSED AND SE
EK WAYS TO APPLY THE NEW CONCEPTS TO THEIR APPLICATIONS WUT
THEN DISILLUSIONMENT SETS IN SYSTEM CONYERSION AND CUSTOMIZA
TION ALWAYS PROYES TO WE A WIGGER XOW THAN ANYONE INCLUDING
MIS COULD HAYE PREDICTED COSTS SOAR YENDORS PRODUCTS NEYER S
EEM TO ZUITE LIYE UP TO THE PROMISES USERS REYERT TO THE TRI
ED AND TRUE AND DECIDE TO LEAYE THE RISKS TO THE PIONEERS TH
IS SEEMS TO WE THE CASE WITH AT LEAST THREE OF THE HIGHER PR
OFILE STANDARDSMAKING EFFORTS ELECTRONIC DATA INTERCHANGE ED
I SEE STORY P MANUFACTURING AUTOMATION PROTOCOLTECHNICAL OFF
ICE PROTOCOL MAPTOP SEE STORY P AND OPEN SYSTEMS INTERCONNEC
TION OSI SEE STORY P IN EACH CASE THE MOYEMENT NEYER SPREAD
ZUITE AS ZUICKLY AS FIRST ANTICIPATED EDI WHICH PROPOSES TO
MOYE WUSINESS DOCUMENTS SUCH AS INYOICES AND PURCHASE ORDERS
 ELECTRONICALLY WETWEEN COMPANIES PROYED SLOWER AND MORE EXP
ENSIYE THAN EXPECTED EARLY ENTHUSIASM FOR MAP STANDARDIZED C
OMMUNICATIONS FOR DEYICES ON THE FACTORY FLOOR HAS GIYEN WAY
 TO SKEPTICISM MOSTLY WECAUSE OF THE SLOW DEYELOPMENT OF MAP
 STANDARDS AND PRODUCTS AND OSI A GLOWAL EFFORT TO ACHIEYE I
NTEROPERAWILITY WETWEEN A WIDE RANGE OF COMPUTER SYSTEMS FAC
ES NATIONAL CONFLICTS WUT IS THAT ANY REASON FOR USERS TO FR
ET AT THE SLOW PACE ACCEPTANCE OF STANDARDS NOT AT ALL STAND
ARDS REMAIN A NECESSARY PART IN THE EYOLUTION OF SYSTEMS REG
ARDLESS OF HOW SLOWLY THE FIRST STEPS ARE TAKEN THE EXPERTS
HAYE REYISED THEIR FORECASTS AND WHILE ACKNOWLEDGING THAT AC
CEPTANCE IS COMING SLOWER THAN FIRST ANTICIPATED IT IS WORTH
 NOTING THAT THEIR OPTIMISM REMAINS SURYEYS INDICATE TOO THA
T USERS INTEREST IN STANDARDS IS WELL ESTAWLISHED ITS XUST T
HAT MOST OF THEM HAYE DECIDED TO WAIT FOR EXAMPLE WHILE MUCH
 OF THE WHOLEHEARTED ENTHUSIASM IS GONE AND SOME ZUESTIONED
THE YIAWILITY OF MAP PRODUCTS THE UNDERLYING NEED FOR AN IND
USTRY STANDARD ON THE FACTORY FLOOR REMAINS SOME OF THE PROW
LEM WILL WE SOLYED IN TIME THROUGH EDUCATION OSI WATCHERS IN
 THE FEDERAL GOYERNMENT NOTE DIFFERENT LEYELS OF KNOWLEDGE A
MONG DIFFERENT USERS THEIR ADYICE IS THAT FOR NOW USERS STAY
 UP TO DATE WITH OSI DEYELOPMENTS LITERATURE AND SO ON OR TH
EY CAN FIND THEMSELYES ZUICKLY FALLING WEHIND FOR ALL THE EX
PECTED SHORT TERM PROWLEMS AND COSTS ASSOCIATED WITH THE EAR
LY IMPLEMENTATION OF STANDARDS THE LONG TERM GAINS STILL MAK
E IT ALL WORTHWHILE REDUCED RESEARCH AND DEYELOPMENT COSTS F
EWER GATEWAYS AND THEIR ASSOCIATED COSTS AND REDUCED PERFORM
ANCE AND CONFORMANCE TESTING
```

Figure 5.9: Cryptanalysis of polyalphabetic substitution using linear discriminant, 20 variables, example 2

```
123123123123123123123123123123123123123123123123123123123123123

EDITORIOL STRNCRRCS UNFOLD RS THEY SHOULD INVRRIRBLY STONFOR
FS MO'EUENTS SEEU TO GROF KY LEOPS IND WOUNDS IT IS OLL TOO
CREFILTOKLE IN THE EIRLY STRBES I FEF VISIONIRIES SCRERD THE
 UESSIGE FITH IN EYONGELIAIL JEOL USERS ORE IMPRESSED RNC SE
EK FOYS TO ICPLY THE NEB AONAECTS TO THEIR OPCLICOTIONS KUT
THEN CISILLUSIONUENT SETS IN SYSTEU CONYERSION IND CUSTOPIXI
TION ILGIYS CROYES TO BE O WIGGER JOW THON ONYONE INALUDINB
MIS AOULC HIYE CREFILTEC AOSTS SORR YENFORS PROCUATS NEVER S
EEP TO XUITE LIYE UP TO THE PROUISES USERS REYERT TO THE TRI
ED RNC TRUE RNC DECICE TO LEOVE THE RISKS TO THE PIONEERS TH
IS SEEMS TO WE THE AISE FITH IT LEOST THREE OG THE HIGHER PR
OFILE STINDIRDSMOKING EFGORTS ELECTRONIC CRTI INTERCHINBE EC
I SEE STORV C PRNUGOLTURING OUTOPRTION CROTOLOLTEAHNILRL OFF
IAE PROTOCOL UICTOC SEE STORV C IND OPEN SYSTEUS INTERCONNEL
TION OSI SEE STORY P IN EOLH LRSE THE POVEMENT NEVER SCRERD
QUITE RS QUICKLY IS FIRST INTICIPRTEF EFI FHICH CROCOSES TO
MOVE KUSINESS COAUMENTS SULH IS INVOIAES IND CURCHISE ORCERS
 ELEATRONIAILLY BETGEEN COPCONIES PROVEF SLOBER RNC UORE EXP
ENSIVE THRN EXPEATED EORLV ENTHUSIOSM FOR MOP STRNCRRCIXEF L
OUPUNICOTIONS FOR FEVIAES ON THE GOLTORY FLOOR HRS GIVEN FOY
 TO SKEPTIAISU MOSTLY BECOUSE OF THE SLOG FEVELOCUENT OF MOP
 STRNCRRCS IND CROFULTS RNC OSI I BLOBIL EGFORT TO RAHIEVE I
NTEROPERIWILITY BETGEEN R BIDE RINBE OF AOMPUTER SYSTEMS GOL
ES NOTIONRL CONGLICTS BUT IS THOT ONY REOSON GOR USERS TO FR
ET RT THE SLOG COLE ICAECTINAE OF STRNCRRCS NOT RT RLL STRNC
RRCS REUIIN R NEAESSIRV CORT IN THE EVOLUTION OF SYSTEMS REG
RRCLESS OG HOG SLOFLY THE FIRST STEPS IRE TOKEN THE EXCERTS
HOVE REVISEC THEIR FORELRSTS IND FHILE RAKNOBLECGING THOT OL
CEPTONCE IS COPING SLOGER THON FIRST ONTILIPITEC IT IS GORTH
 NOTING THOT THEIR OPTIPISP REMOINS SURYEYS INDICOTE TOO THI
T USERS INTEREST IN STONFORFS IS FELL ESTIWLISHEF ITS JUST T
HOT UOST OF THEM HRVE DECICED TO FOIT FOR EXIMPLE BHILE PUAH
 OF THE BHOLEHERRTED ENTHUSIOSM IS GONE ONF SOUE XUESTIONED
THE VIRBILITY OG PRP CROFULTS THE UNFERLVINB NEEF FOR RN INC
USTRV STINDIRD ON THE FICTORV GLOOR REPRINS SOUE OF THE PROK
LEP GILL WE SOLYEC IN TIME THROUGH EFULRTION OSI GITAHERS IN
 THE FEDEROL BOYERNUENT NOTE CIFFERENT LEVELS OF KNOGLEDGE I
MONG CIFFERENT USERS THEIR RDVIAE IS THRT GOR NOF USERS STOY
 UP TO DITE FITH OSI FEVELOCUENTS LITERRTURE RNC SO ON OR TH
EV CON FIND THEMSELVES AUILKLY FILLINB WEHINC FOR ILL THE EX
CELTEC SHORT TERU CROWLEMS RNC AOSTS OSSOLIOTED FITH THE EOR
LV IUPLEPENTRTION OF STINDIRDS THE LONG TERU GOINS STILL MOK
E IT ILL FORTHBHILE REDUCEC RESEIRAH ONF CEVELOPMENT LOSTS F
EGER GRTEFOYS IND THEIR RSSOAIRTEF LOSTS IND RECUAEF PERFORP
RNLE IND CONGORMONCE TESTINTI
```

Figure 5.10: Cryptanalysis of polyalphabetic substitution using quadratic discriminant, 20 variables, example 2

```
123123123123123123123123123123123123123123123123123123123123123

EDITOROAL STTSDTRDS UNFOLD TS THEY SHOULD OSHTRITGLY STANHAR
HS MOHECENNS SEEC TO GROH KY LEAUS RND WOUNDS IT IS ALL TOO
CREHILTAKLE OS THE ERCLH STTBES R FEH HOSIOSRCIES SCRETD THE
 CESSRGE HITH RN EYANGELOLRL ZEAL USERS ARE IMPRESSED TSD SE
EK HAHS TO RCPLY THE NEG LONLECNS NO NHEIR AUCLILATOONS KUN
THEN DOSILLUSIONCENN SETS IN SYSTEC LONYERSION RND LUSTOPOXR
TION RLPRYS CROYES NO GE A WIGGER ZOW THAN ANYONE INLLUDINB
MIS LOULD HRYF CREHILTED LOSNS SOTR YENHORS UCODULTS NEVEC S
EEP NO XUONE LIYE UP TO THE PROCISES USERS REYERT TO THE TRI
ED TSD NRUE TSD DELIDE TO LEAHE THE CISKS TO THE PIOSEERS NH
OS SEEMS TO WE THE LRSE HITH RT LEAST THREE OP THE HIGHER PR
OFILE SNRNDRCDSMAKOSG EHPORTS ELELNROSIL DTNR INTERLHRNBE ED
O SEE SNOCW C PTSUPALTUROSG AUTOPTNIOS CROTOLOLTELHNILTL OFH
OLE PRONOLOL CRCNOC SEE SNOCW C RND OPEN SYSTECS INTERLONNEL
TION OSI SEE STORH P OS EALH LTSE NHE POVEMENT NEVEC SCRETD
XUITE TS XUILKLY RS HORST RNNILIUTNEH EHI HHILH CROCOSES TO
MOHE KUSINESS DOLUMENTS SULH RS INVOOLES RND CURLHRSE ORDERS
 ELELTCONOLRLLH GETPEES LOPCANOES PROVEH SLOGER TSD COCE EXU
ESSOVE NHTS EXUELTED EARLW ESTHUSOASM HOR MAU STTSDTRDOXEH L
OCPUSILATOONS HOP HEHOLES ON THE PALTORY HLOOC HTS GIHES HAH
 NO SKEPTOLISC MOSTLH GELAUSE OF THE SLOP HEHELOCCENN OF MAU
 STTSDTRDS RND CROHULTS TSD OSO R BLOGRL EPFOCN TO TLHOEHE I
NNECOUERRWILONH GETPEES T GODE RRNBE OH LOMPUTER SHSNEMS PAL
ES NATOONTL LONPLIINS GUT IS THAT ANY REASOS POR USERS NO FR
EN TN THE SLOP CALE RLLECNRNLE OH STTSDTRDS NON TN TLL STTSD
TRDS RECROS T NELESSRCW CART IN THE EVOLUTOON OH SHSNEMS CEG
TRDLESS OP HOP SLOHLH NHE HORST STEUS RCE TAKES THE EKCERTS
HAHE REVISED NHEIR FOCELTSTS RND HHILE TLKNOGLEDGING THAT AL
LEUTANLE OS LOPOSG SLOPEC THAN FICST ANTILOPRTED IT IS POCNH
 SOTING THAT NHEIR OUTIPOSP REMAINS SURYEHS INDILATE TOO THR
T USERS INNECEST IN STANHARHS OS HELL ESNRWLISHEH ITS ZUST T
HAT COSN OF THEM HTVE DELIDED TO HAIT HOR EXRMPLE GHILE PULH
 OH NHE GHOLEHETRTED ESTHUSOASM IS GOSE ANH SOCE XUESTUONED
THE VITGILITY OP PTP CROHULTS THE UNHERLWINB NEEH HOR TS OSD
USTCW SNRNDRCD OS THE FRLNOCW PLOOR CEPTINS SOCE OH NHE UCOK
LEP PILL WE SOLYED IN NIME THROUGH EHULTNIOS OSI PRTLHERS IN
 NHE HEDECAL BOYERNCENN NOTE DOFHERENN LEHELS OH KNOPLEDGE R
MONG DOFHERENN USECS THEOR TDHOLE IS NHTN POR SOH USERS STAH
 UU NO DRTE HITH OSI HEHELOCCENNS LITERTNUCE TSD SO ON OR NH
EW LAN FIND THEMSELVES ZUILKLH FRLLINB WEHOSD FOC RLL THE EK
CELTED SHORT NECC CROWLEMS TSD LOSNS ASSOLOATED HITH THE EAR
LW OCULEPESTTNIOS OF SNRNDRCDS NHE LOSG NECC GAINS SNILL MAK
E IT RLL HORTHGHILE REDULED RESERCLH ANH DEVELOUMENT LOSTS H
EPEC GTNEHAHS RND THEOR TSSOLITNEH LOSTS RND CEDULEH UERHORI
TSLE RND LONPORMANLE TESTINTI
```

CHAPTER 6

CONCLUSIONS AND SUGGESTION

The objective of our research is to extract from an encrypted document all the information we can relevant to the true representation of its symbols in a plain or source text alphabet. In summary, our conclusions are methodological, prescriptive, and suggestive.

6.1 Methodological point of view

Our methodology is :

1. Select two random samples of plaintext similar to that which we will cryptanalyze. One sample is called the "training sample" ; the other is called the "validating sample".
2. The set of all occurences of a character type in the plaintext alphabet is called a "group". Each group has a vector of measurements called a "measurement vector".
3. The components of the measurement vector include the relative frequency and other variables that can be generated from one-graph and digraph structures of the text representing information measures from one and two-dimensional probability spaces, respectively.
4. A linear discriminant function, and a quadratic discriminant function are computed for each group. The training sample is used to generate the coefficients and constants of the measurement vector for every group in linear discriminant analysis. The application of the functions to a measurement vector is used to determine into which group the vector should be placed.

We could also use higher dimensional probability spaces to acquire more information but at an immense cost in terms of sample size, computation time and computer memory.

6.2 Prescriptive point of view

From a prescriptive point of view :

(1) This research suggests that stepwise linear discriminant analysis is powerful enough to be used in pattern recognition problems. One does not need to use quadratic discriminant method, which is based on the assumption that some covariance matrices of the groups are not equal. In relation to linear discriminant analysis, James [1985] suggested that pooled sample covariance matrix can be considered as "the average" of different covariance matrices of all groups. In other words, the pooled sample covariance matrix can be justified to be used as the representation of all different covariance matrices and to fulfill the required assumption of equal covariance matrices which leads to the linear discriminant method. The other drawbacks of the quadratic discriminant method is that it requires more computer memory to store all the discriminant function.: of the groups as matrices or two-dimensional arrays. The calculation of discriminant scores will also use more computer processing time. The linear discriminant computation uses only one-dimensional arrays.

(2) The stepwise linear discriminant analysis is useful for selecting variables which make a significant contribution to the building of discriminant functions for all groups of characters.

(3) "Invariant variables" are mathematically derived quantities that are exactly the same for the characters in source and encrypted text, only their arrangement is different. The digraph structure of text files provides more variation or information through the generation of more invariant variables which relate source text and the corresponding crypto text. This fact can be observed in the significant improvement of the percentage of correct classification in monoalphabetic and polyalphabetic substitutions, both in linear and quadratic discriminant analysis. The improvement of total percentage of correctness in monoalphabetic substitution is greater than in polyalphabetic substitution. This is to be expected, since polyalphabetic substitution has more complicated structure than monoalphabetic substitution.

(4) To improve the discriminating power of discriminant analysis, the digraph structure should be extended to trigraph structure in the same way as the one-graph structure was expanded to digraph structure. Trigraph or higher structure will generate more invariant

variables than digraph structure, and this means that one will gain more variation or information to be used for classification. In fact, information provided by variables in the lower structures can also be derived from the higher structures. For instance, the marginal quantities in digraph structures are equivalent to one-graph structures (see Table 2.1 on page 22). The problem with the addition of variables is that the time and space for accomplishing computation will also increase.

6.3 Suggestive point of view

From a suggestive point of view :

The objective of this work has been to explore the limits of multiple discriminant analysis in extracting latent meaning from abstract patterns. We chose as a "test bed" the cryptanalysis of polyalphabetic ciphers. When the characters of plaintext are scrambled by successive application of three or more alphabets, it produces as complex an abstract pattern as one could want ; and a pattern whose meaning is known with certainty.

We accept that few persons desiring communications privacy would use a polyalphabetic cipher today when block-product and exponentiating ciphers are available that provide much greater resistance to cryptanalysis. Computer solution of polyalphabetic ciphers without human intervention or without convergent retries, remains an interesting task in which to study techniques for pattern perception.

Computer analysis of text is a case of conjuring with a longer plaintext alphabet, one with 40,000 symbols instead of 27 or 255 (if one chooses to use the entire ASCII set). Experiments in automatic abstracting, extracting, and indexing based upon the occurences and co-occurences of words have shown some promise but have been called into question by persons who cavil about : "What is the meaning of meaning". Here, on an abstract level it is possible to explore technique where there is no opportunity to quibble about results.

Essentially, anything that can be measured can be expressed in a set of symbols denoting levels of value. This include characters, words, spots of colour, brush strokes, scratches, errors, and all physical, mental, and biological attributes. The co-occurences of these symbols is a pattern whether the occurences are over time or space. The entire universe is made up of patterns. Sometimes it is easy to perceive their meaning. More often the meaning is hidden. All mysteries can be regarded as a code. What we have presented here is a way to structure what we know or can learn from observation so that powerful statistical techniques and computational procedures can help us unravel them.

These techniques can be applied in other fields than cryptanalysis and information science. We have already discussed classification of varieties of iris flowers and optical patterns of printed characters. The latter can, of course, be extended to handwriting analysis and identification of persons.

Hand [1981] suggests these additional applications :

(a) classification of archaeological specimens, for example, a skull to one of two races: English or Eskimo ;

(b) classification of crops from high altitude photographs, which might be cheaper in estimating total acreage than by ground measurement; the same method have also been applied to the detection of mineral deposits ;

(c) speech recognition where the objects to be classified are waveforms, cardiac wave analysis, target recognition from vehicle noise or radar returns, and automatic electro-encephalograph analysis;

(d) recognizing high-risk individuals in psychiatric research.

Lachenbruch [1975] mentions other areas of application which include anthropology, ecology, taxonomy, author verification, psychology, fingerprint recognition, and so on.

There may be other data structures than the linear array. There may be other tools than multivariate discriminant analysis. Some codes in nature and in intellectual products will yield to this technique. Some will yield to other techniques ; or to a combination. But the power of the computer now makes it possible to try them all and perhaps invent others. Ultimately most mysteries will be solved.

APPENDIX A

Results from Linear Discriminant Analysis

Table A.1: Coefficients of linear discrimination for character groups, monoalphabetic, 1 variable

Group	Coefficients of RELFREQ	CONSTANT
A	2451.34766	-88.34501
B	427.62061	-7.06080
C	1247.35107	-25.56044
D	1107.85669	-20.95539
E	3437.08569	-170.82770
F	642.20062	-9.92955
G	527.38367	-8.08854
H	1000.28503	-17.81850
I	2162.29590	-69.45689
J	37.09273	-6.77037
K	187.06538	-5.72156
L	1156.78723	-22.48681
M	931.46771	-16.01187
N	2227.84033	-73.54531
O	2281.71167	-76.98656
P	835.21796	-13.70173
Q	47.97626	-6.80693
R	1933.44788	-56.22342
S	2099.72778	-65.69437
T	2612.28491	-99.92624
U	807.90601	-13.06809
V	322.35156	-6.15932
W	427.76773	-6.98903
X	109.97643	-6.08146
Y	442.89038	-7.12157
Z	23.69461	-7.04929
@	5575.79346	-445.24588

Table A.2: Classification matrix for training sample : monoalphabetic 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.63	31	0	0	0	2	0	0	0	5	0	0	0	0	5
B	0.00	0	0	0	0	0	5	48	0	0	0	0	0	0	0
C	50.50	0	0	51	7	0	0	0	9	0	0	0	17	5	0
D	20.21	0	0	24	19	0	0	0	15	0	0	0	22	8	0
E	92.66	0	0	0	0	101	0	0	0	0	0	0	0	0	0
F	34.48	0	0	0	0	0	30	20	1	0	0	0	0	3	0
G	49.48	0	0	0	0	0	23	48	1	0	0	0	0	1	0
H	19.39	0	0	20	13	0	2	1	19	0	0	0	11	14	0
I	11.36	8	0	0	0	0	0	0	0	10	0	0	0	0	9
J	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
K	13.19	0	0	0	0	0	0	1	0	0	0	12	0	0	0
L	24.24	0	0	33	19	0	0	0	14	0	0	0	24	2	0
M	12.77	0	0	11	6	0	7	1	17	0	0	0	5	12	0
N	11.43	19	0	0	0	0	0	0	0	8	0	0	0	0	12
O	19.05	26	0	1	0	0	0	0	0	9	0	0	0	0	10
P	5.88	0	0	3	5	0	5	5	13	0	0	0	3	22	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	96	0	0	0
R	62.63	0	0	5	0	0	0	0	0	9	0	0	0	0	7
S	11.88	12	0	4	0	0	0	0	0	12	0	0	0	0	8
T	52.75	22	0	0	0	9	0	0	0	3	0	0	0	0	2
U	30.63	0	0	1	3	0	13	6	15	0	0	0	1	24	0
V	30.77	0	0	0	0	0	2	19	0	0	0	1	0	0	0
W	4.30	0	0	0	0	0	6	40	0	0	0	0	0	1	0
X	0.00	0	0	0	0	0	0	1	0	0	0	63	0	0	0
Y	28.71	0	0	0	0	0	12	49	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	101	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	26.81	118	0	153	72	112	105	239	104	56	0	369	83	92	53

Table A.2 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	31.63	19	0	0	1	5	30	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	2	8	2	0	31	0	0	96
C	50.50	0	2	0	9	0	0	1	0	0	0	0	0	0	101
D	20.21	0	1	0	2	0	0	3	0	0	0	0	0	0	94
E	92.66	0	0	0	0	0	8	0	0	0	0	0	0	0	109
F	34.48	0	1	0	0	0	0	31	0	0	0	1	0	0	87
G	49.48	0	0	0	0	0	0	10	2	0	0	12	0	0	97
H	19.39	0	3	0	0	0	0	14	0	1	0	0	0	0	98
I	11.36	16	0	0	20	17	8	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	4	0	0	0	0	0	100
K	13.19	0	0	0	0	0	0	1	69	3	0	5	0	0	91
L	24.24	0	1	0	4	0	0	2	0	0	0	0	0	0	99
M	12.77	0	11	0	2	0	0	22	0	0	0	0	0	0	94
N	11.43	19	0	0	19	17	11	0	0	0	0	0	0	0	105
O	19.05	20	0	0	12	11	16	0	0	0	0	0	0	0	105
P	5.88	0	6	0	0	0	0	40	0	0	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	6	0	0	0	0	0	102
R	62.63	3	0	0	62	12	1	0	0	0	0	0	0	0	99
S	11.88	11	0	0	34	12	8	0	6	0	0	0	0	0	101
T	52.75	4	0	0	2	1	48	0	0	0	0	0	0	0	91
U	30.63	0	14	0	0	0	0	34	0	0	0	0	0	0	111
V	30.77	0	0	0	0	0	0	0	32	14	0	36	0	0	104
W	4.30	0	0	0	0	0	0	3	7	4	0	32	0	0	93
X	0.00	0	0	0	0	0	0	0	40	1	0	0	0	0	105
Y	28.71	0	0	0	0	0	0	2	3	6	0	29	0	0	101
Z	0.00	0	0	0	0	0	0	0	1	0	0	0	0	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	26.81	92	39	0	167	75	130	165	172	31	0	146	0	101	2674

Table A.3: Classification matrix for validating sample : monoalphabetic 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	31.37	32	0	0	0	1	0	0	0	6	0	0	0	0	5
B	0.00	0	0	0	0	0	6	40	0	0	0	0	0	1	0
C	47.47	0	0	47	11	0	0	0	10	0	0	0	13	9	0
D	13.21	0	0	37	14	0	0	0	20	0	0	0	20	9	0
E	93.41	0	0	0	0	85	0	0	0	0	0	0	0	0	0
F	33.63	0	0	0	2	0	38	30	1	0	0	0	0	6	0
G	47.57	0	0	0	0	0	26	49	0	0	0	0	0	1	0
H	12.75	0	0	18	15	0	2	0	13	0	0	0	20	15	0
I	12.50	15	0	0	0	0	0	0	0	14	0	0	0	0	8
J	0.00	0	0	0	0	0	0	0	0	0	0	98	0	0	0
K	22.94	0	0	0	0	0	0	1	0	0	0	25	0	0	0
L	17.82	0	0	38	17	0	0	0	14	0	0	0	18	7	0
M	20.75	0	0	5	13	0	5	1	19	0	0	0	9	22	0
N	8.42	13	0	0	0	1	0	0	0	15	0	0	0	0	8
O	23.16	20	0	0	0	0	0	0	0	12	0	0	0	0	15
P	9.18	0	0	3	5	0	11	5	8	0	0	0	1	18	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	90	0	0	0
R	51.49	1	0	4	0	0	0	0	0	8	0	0	0	0	8
S	14.14	12	0	0	0	0	0	0	0	13	0	0	0	0	8
T	68.81	17	0	0	0	6	0	0	0	1	0	0	0	0	1
U	34.83	0	0	0	2	0	7	5	17	0	0	0	4	18	0
V	47.92	0	0	0	0	0	0	7	0	0	0	1	0	0	0
W	10.28	0	0	0	0	0	14	45	0	0	0	0	0	0	0
X	0.00	0	0	0	0	0	0	0	0	0	0	65	0	0	0
Y	26.26	0	0	0	0	0	9	53	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	97	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	27.51	110	0	152	79	93	118	236	102	69	0	376	85	106	53

Table A.3 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	31.37	14	0	0	0	2	42	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	4	8	4	0	41	0	0	104
C	47.47	0	1	0	5	0	0	3	0	0	0	0	0	0	99
D	13.21	0	0	0	1	0	0	5	0	0	0	0	0	0	106
E	93.41	0	0	0	0	0	6	0	0	0	0	0	0	0	91
F	33.63	0	0	0	0	0	0	33	0	0	0	3	0	0	113
G	47.57	0	1	0	0	0	0	14	0	0	0	12	0	0	103
H	12.75	0	2	0	1	0	0	16	0	0	0	0	0	0	102
I	12.50	18	0	0	33	15	9	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	2	0	0	0	0	0	100
K	22.94	0	0	0	0	0	0	0	67	10	0	6	0	0	109
L	17.82	0	1	0	4	0	0	2	0	0	0	0	0	0	101
M	20.75	0	11	0	0	0	0	21	0	0	0	0	0	0	106
N	8.42	7	0	0	28	14	9	0	0	0	0	0	0	0	95
O	23.16	22	0	0	6	10	10	0	0	0	0	0	0	0	95
P	9.18	0	9	0	0	0	0	38	0	0	0	0	0	0	98
Q	0.00	0	0	0	0	0	0	0	8	0	0	0	0	0	98
R	51.49	8	0	0	52	20	0	0	0	0	0	0	0	0	101
S	14.14	24	0	0	26	14	2	0	0	0	0	0	0	0	99
T	68.81	7	0	0	1	1	75	0	0	0	0	0	0	0	109
U	34.83	0	5	0	0	0	0	31	0	0	0	0	0	0	89
V	47.92	0	0	0	0	0	0	0	46	10	0	32	0	0	96
W	10.28	0	0	0	0	0	0	2	7	11	0	28	0	0	107
X	0.00	0	0	0	0	0	0	0	29	0	0	1	0	0	95
Y	26.26	0	0	0	0	0	0	3	5	3	0	26	0	0	99
Z	0.00	0	0	0	0	0	0	0	0	1	0	0	0	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	27.51	100	30	0	157	76	153	172	172	39	0	149	0	99	2726

Table A.4: Coefficients of linear discriminant functions for character groups : monoalphabetic, 3 variables

No.	Group	RELFREQ	INFOCONT	CTOENTRO	CONSTANT
1	A	-3687.79395	7.67708	3179.47559	-179.12953
2	B	-3457.97314	7.63469	2048.10645	-54.98252
3	C	-5006.87598	7.42002	3233.91431	-117.77074
4	D	-4964.45605	7.39953	3142.36646	-108.81699
5	E	-829.49176	7.84456	2241.98584	-225.67661
6	F	-4211.05859	7.44895	2531.28564	-72.84222
7	G	-3833.99072	7.53115	2285.51807	-62.86501
8	H	-4822.03613	7.40441	3016.98438	-100.07224
9	I	-4262.04297	7.61352	3321.76782	-166.70990
10	J	-689.60516	9.37557	485.40363	-37.96316
11	K	-2090.75903	8.35775	1250.64539	-39.07847
12	L	-4990.19775	7.40505	3179.90771	-112.10403
13	M	-4728.29639	7.40294	2935.39453	-94.72527
14	N	-4151.44727	7.62853	3299.35449	-169.78030
15	O	-4037.22046	7.64041	3269.22168	-171.82904
16	P	-4620.40625	7.40268	2832.95459	-88.12072
17	Q	-807.27124	8.99306	544.98297	-35.91635
18	R	-4628.45654	7.56081	3390.11890	-156.60033
19	S	-4340.36084	7.59895	3329.48340	-163.25862
20	T	-3298.38379	7.70969	3065.24854	-185.66403
21	U	-4581.15674	7.40481	2799.58179	-86.13235
22	V	-2952.74609	7.82639	1744.23047	-46.88622
23	W	-3447.56836	7.63813	2043.00342	-54.77885
24	X	-1460.68335	9.03790	904.55212	-38.52266
25	Y	-3518.67432	7.61785	2086.01147	-56.07713
26	Z	-506.77835	8.67619	377.93115	-33.48726
27	@	8720.47656	7.88417	-1476.45093	-486.86832

Table A.5: Classification matrix for training sample : monoalphabetic 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	30.61	30	0	0	0	1	0	0	0	5	0	0	0	0	6
B	0.00	0	0	0	0	0	6	31	0	0	0	2	0	0	0
C	52.48	0	0	53	7	0	0	0	9	0	0	0	15	5	0
D	20.21	0	0	26	19	0	0	0	15	0	0	0	20	7	0
E	92.66	0	0	0	0	101	0	0	0	0	0	0	0	0	0
F	42.53	0	0	0	0	0	37	20	1	0	0	0	0	2	0
G	38.14	0	0	0	0	0	26	37	1	0	0	0	0	1	0
H	18.37	0	0	19	13	0	3	1	18	0	0	0	11	14	0
I	11.36	9	0	0	0	0	0	0	0	10	0	0	0	0	11
J	70.00	0	0	0	0	0	0	0	0	0	70	4	0	0	0
K	62.64	0	0	0	0	0	0	0	0	0	3	57	0	0	0
L	23.23	0	0	34	19	0	0	0	14	0	0	0	23	2	0
M	13.83	0	0	11	7	0	8	1	15	0	0	0	5	13	0
N	10.48	19	0	0	0	0	0	0	0	9	0	0	0	0	11
O	19.05	26	0	1	0	0	0	0	0	9	0	0	0	0	11
P	6.86	0	0	3	5	0	12	5	13	0	0	0	3	21	0
Q	0.00	0	0	0	0	0	0	0	0	0	58	4	0	0	0
R	65.66	0	0	2	0	0	0	0	0	9	0	0	0	0	8
S	10.89	12	0	4	0	0	0	0	0	12	0	0	0	0	8
T	53.85	23	0	0	0	7	0	0	0	3	0	0	0	0	2
U	27.03	0	0	1	3	0	14	6	15	0	0	0	1	23	0
V	53.85	0	0	0	0	0	2	7	0	0	0	10	0	0	0
W	24.73	0	0	0	0	0	6	23	0	0	0	0	0	1	0
X	42.86	0	0	0	0	0	0	0	0	0	19	33	0	0	0
Y	17.82	0	0	0	0	0	12	32	0	0	0	2	0	0	0
Z	18.63	0	0	0	0	0	0	0	0	0	74	1	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	34.52	119	0	154	73	109	126	163	101	57	224	113	78	89	57

Table A.5 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	30.61	21	0	0	1	5	29	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	1	16	23	0	17	0	0	96
C	52.48	0	2	0	9	0	0	1	0	0	0	0	0	0	101
D	20.21	0	2	0	2	0	0	3	0	0	0	0	0	0	94
E	92.66	0	0	0	0	0	8	0	0	0	0	0	0	0	109
F	42.53	0	2	0	0	0	0	22	0	1	0	2	0	0	87
G	38.14	0	0	0	0	0	0	6	7	7	0	12	0	0	97
H	18.37	0	5	0	1	0	0	12	1	0	0	0	0	0	98
I	11.36	14	0	0	21	16	7	0	0	0	0	0	0	0	88
J	70.00	0	0	0	0	0	0	0	0	0	17	0	9	0	100
K	62.64	0	0	0	0	0	0	1	18	2	9	1	0	0	91
L	23.23	0	1	0	4	0	0	2	0	0	0	0	0	0	99
M	13.83	0	12	0	2	0	0	20	0	0	0	0	0	0	94
N	10.48	19	0	0	19	17	11	0	0	0	0	0	0	0	105
O	19.05	20	0	0	12	11	15	0	0	0	0	0	0	0	105
P	6.86	0	7	0	0	0	0	33	0	0	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	2	0	28	0	10	0	102
R	65.66	2	0	0	65	12	1	0	0	0	0	0	0	0	99
S	10.89	10	0	0	36	11	8	0	0	0	0	0	0	0	101
T	53.85	4	0	0	2	1	49	0	0	0	0	0	0	0	91
U	27.03	0	17	0	0	0	0	30	0	0	0	1	0	0	111
V	53.85	0	0	0	0	0	0	0	56	16	1	12	0	0	104
W	24.73	0	0	0	0	0	0	3	21	23	0	16	0	0	93
X	42.86	0	0	0	0	0	0	0	7	0	45	1	0	0	105
Y	17.82	0	0	0	0	0	0	1	20	16	0	18	0	0	101
Z	18.63	0	0	0	0	0	0	0	0	0	8	0	19	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	34.52	90	48	0	174	73	128	135	148	88	108	80	38	101	2674

Table A.6: Classification matrix for validating sample : monoalphabetic 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	27.45	28	0	0	0	1	0	0	0	7	0	0	0	0	4
B	0.00	0	0	0	0	0	6	34	0	0	0	3	0	1	0
C	47.47	0	0	47	11	0	0	0	10	0	0	0	13	7	0
D	14.15	0	0	37	15	0	0	0	20	0	0	0	19	8	0
E	91.21	0	0	0	0	83	0	0	0	0	0	0	0	0	0
F	41.59	0	0	0	2	0	47	27	1	0	0	0	0	5	0
G	33.98	0	0	0	0	0	29	35	0	0	0	0	0	1	0
H	12.75	0	0	18	15	0	4	0	13	0	0	0	20	15	0
I	12.50	17	0	0	0	0	0	0	0	14	0	0	0	0	9
J	62.00	0	0	0	0	0	0	0	0	0	62	2	0	0	0
K	46.79	0	0	0	0	0	0	1	0	0	2	51	0	0	0
L	15.84	0	0	39	18	0	0	0	13	0	0	0	16	7	0
M	20.75	0	0	5	14	0	8	1	19	0	0	0	8	22	0
N	9.47	13	0	0	0	1	0	0	0	15	0	0	0	0	9
O	22.11	21	0	0	0	0	0	0	0	11	0	0	0	0	16
P	12.24	0	0	3	5	0	18	5	8	0	0	0	1	16	0
Q	0.00	0	0	0	0	0	0	0	0	0	58	7	0	0	0
R	54.46	1	0	4	0	0	0	0	0	9	0	0	0	0	8
S	15.15	11	0	0	0	0	0	0	0	12	0	0	0	0	9
T	68.81	17	0	0	0	5	0	0	0	2	0	0	0	0	0
U	31.46	0	0	0	2	0	10	5	17	0	0	0	4	17	0
V	47.92	0	0	0	0	0	0	3	0	0	0	22	0	0	0
W	13.08	0	0	0	0	0	13	28	0	0	0	1	0	0	0
X	48.42	0	0	0	0	0	0	0	0	0	18	27	0	0	0
Y	23.23	0	0	0	0	0	7	32	0	0	0	1	0	0	0
Z	14.29	0	0	0	0	0	0	0	0	0	72	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	32.65	108	0	153	82	90	142	171	101	70	212	114	81	99	55

Table A.6 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	27.45	18	0	0	0	2	42	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	3	25	25	0	7	0	0	104
C	47.47	0	3	0	5	0	0	3	0	0	0	0	0	0	99
D	14.15	0	1	0	1	0	0	5	0	0	0	0	0	0	106
E	91.21	0	0	0	0	0	8	0	0	0	0	0	0	0	91
F	41.59	0	1	0	0	0	0	23	1	2	0	4	0	0	113
G	33.98	0	1	0	0	0	0	10	4	8	0	15	0	0	103
H	12.75	0	2	0	1	0	0	14	0	0	0	0	0	0	102
I	12.50	17	0	0	33	15	7	0	0	0	0	0	0	0	112
J	62.00	0	0	0	0	0	0	0	0	0	28	0	8	0	100
K	46.79	0	0	0	0	0	0	0	30	2	23	0	0	0	109
L	15.84	0	1	0	5	0	0	2	0	0	0	0	0	0	101
M	20.75	0	12	0	0	0	0	17	0	0	0	0	0	0	106
N	9.47	6	0	0	28	14	9	0	0	0	0	0	0	0	95
O	22.11	21	0	0	6	11	9	0	0	0	0	0	0	0	95
P	12.24	0	12	0	0	0	0	29	0	0	0	1	0	0	98
Q	0.00	0	0	0	0	0	0	0	1	0	23	0	9	0	98
R	54.46	7	0	0	55	17	0	0	0	0	0	0	0	0	101
S	15.15	25	0	0	26	15	1	0	0	0	0	0	0	0	99
T	68.81	8	0	0	1	1	75	0	0	0	0	0	0	0	109
U	31.46	0	6	0	0	0	0	28	0	0	0	0	0	0	89
V	47.92	0	0	0	0	0	0	0	46	20	1	4	0	0	96
W	13.08	0	0	0	0	0	0	2	32	14	0	17	0	0	107
X	48.42	0	0	0	0	0	0	0	2	1	46	0	1	0	95
Y	23.23	0	0	0	0	0	0	2	17	17	0	23	0	0	99
Z	14.29	0	0	0	0	0	0	0	1	0	11	0	14	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	32.65	102	39	0	161	75	151	138	159	89	132	71	32	99	2726

Table A.7: Coefficients of linear discriminant functions for character groups : monoalphabetic, 12 variables

No.	Variable	A	B	C
1	RELFREQ	-11943.66309	-14761.09473	-16612.90625
2	ROWCOLSS	280143.28125	979428.25000	900528.62500
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-36.20770	-49.04217	-40.80407
6	CWCPROSS	-10.81217	-2.45643	-16.24651
7	INFOCONT	12.49839	12.70577	13.08094
8	CTOENTRO	3520.39819	5630.56689	6355.04102
9	RCTOJENT	546.69806	-9.17379	280.79944
10	CCTOJENT	949.55225	-55.89225	29.60854
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-219.27435	-88.46635	-152.25735

Table A.7 (continued)

No.	Variable	D	E	F
1	RELFREQ	-16537.80469	-9103.12012	-16488.04688
2	ROWCOLSS	990955.37500	172624.34375	1066528.37500
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-32.44326	-40.34113	-42.93111
6	CWCPROSS	-20.20465	-1.25898	-9.71234
7	INFOCONT	12.54381	12.11807	12.59459
8	CTOENTRO	6751.73193	2039.49243	6422.49512
9	RCTOJENT	-293.76947	-153.12364	11.77005
10	CCTOJENT	268.51849	2024.07104	-87.18366
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-142.63394	-296.98483	-110.07606

Table A.7 (continued)

No.	Variable	G	H	I
1	RELFREQ	-15790.19531	-15541.18457	-14035.90918
2	ROWCOLSS	1026203.43750	1036325.81250	396912.37500
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-38.67307	-38.51013	37.36470
6	CWCPROSS	-13.24643	-8.03588	-9.30059
7	INFOCONT	12.60512	11.92393	12.34468
8	CTOENTRO	6013.47070	6974.85498	4059.13623
9	RCTOJENT	-68.21793	-199.74898	269.56491
10	CCTOJENT	37.79459	-242.26340	1270.51880
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-97.36857	-132.80991	-214.53152

Table A.7 (continued)

No.	Variable	J	K	L
1	RELFREQ	-2727.35303	-9351.91504	-17990.80273
2	ROWCOLSS	192893.48438	665802.37500	929195.93750
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-17.50242	-31.04651	-41.64822
6	CWCPROSS	2.31014	-19.90595	-16.91795
7	INFOCONT	10.91374	13.45840	13.19949
8	CTOENTRO	1238.77881	3792.16235	6254.07910
9	RCTOJENT	-38.23935	-163.61668	133.65401
10	CCTOJENT	11.72915	-20.36639	495.97495
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-39.84684	-57.05044	-153.94402

Table A.7 (continued)

No.	Variable	M	N	O
1	RELFREQ	-17305.89258	-9969.45898	-13231.58887
2	ROWCOLSS	1017682.18750	412189.40625	333439.43750
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-42.95556	-32.24918	-34.34837
6	CWCPROSS	-16.11058	-27.68642	-12.93669
7	INFOCONT	13.22969	13.78579	12.46471
8	CTOENTRO	6538.75049	4700.97168	3744.75830
9	RCTOJENT	82.56463	985.47290	633.83612
10	CCTOJENT	103.92649	-696.04193	961.45703
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-133.48476	-200.66878	-216.49864

Table A.7 (continued)

No.	Variable	P	Q	R
1	RELFREQ	-17858.54102	-3278.25415	-13539.06738
2	ROWCOLSS	1033656.56250	211864.87500	532208.31250
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-43.27249	29.26495	-35.37039
6	CWCPROSS	-15.05980	-17.14374	-19.05752
7	INFOCONT	13.14577	7.59678	13.07567
8	CTOENTRO	6415 96240	1236.08923	4965.96582
9	RCTOJENT	172.50098	99.41706	575.28534
10	CCTOJENT	160.47484	-80.67850	225.53587
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-129.61407	-40.38993	-185.29568

Table A.7 (continued)

No.	Variable	S	T	U
1	RELFREQ	-11943.08887	-8240.72168	-18610.71094
2	ROWCOLSS	510204.87500	304554.18750	1025820.18750
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-39.03479	-41.83262	-42.47076
6	CWCPROSS	-13.42318	-18.18671	-18.33711
7	INFOCONT	12.89285	13.90270	13.39701
8	CTOENTRO	5042.36865	4082.89038	6231.70068
9	RCTOJENT	-768.32611	30.08813	121.95776
10	CCTOJENT	1224.48657	353.46112	477.43991
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-199.80290	-203.36232	-133.16563

Table A.7 (continued)

No.	Variable	V	W	X
1	RELFREQ	-12725.52930	-14489.50781	-6585.81104
2	ROWCOLSS	880873.93750	979927.43750	471777.93750
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-10.98488	-48.16063	-41.59956
6	CWCPROSS	-29.56168	-0.27389	12.19488
7	INFOCONT	11.71926	12.38989	11.99098
8	CTOENTRO	5005.47363	5660.98926	2705.64624
9	RCTOJENT	-267.05487	-48.49443	-179.15225
10	CCTOJENT	95.82621	-96.60343	40.28438
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-72.59393	-87.17426	-49.29915

Table A.7 (continued)

No.	Variable	Y	Z	@
1	RELFREQ	-14709.30176	-1906.38416	-21745.68359
2	ROWCOLSS	984270.18750	125962.79688	2036274.12500
3	WINROWSS	0.00000	0.00000	0.00000
4	WINCOLSS	0.00000	0.00000	0.00000
5	RWCPROSS	-6.82877	-10.01246	-14.03663
6	CWCPROSS	-31.86215	5.11044	-22.89501
7	INFOCONT	11.26394	9.15539	10.96030
8	CTOENTRO	5675.18994	846.80872	4633.22070
9	RCTOJENT	-382.33521	-50.97153	2036.60901
10	CCTOJENT	276.90433	25.25360	46.94153
11	RWCENTRO	0.00000	0.00000	0.00000
12	CWCENTRO	0.00000	0.00000	0.00000
	CONSTANT	-88.82454	-34.20631	-678.86218

Table A.8: Classification matrix for training sample : monoalphabetic 12 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	54.08	53	0	0	0	1	0	0	0	10	0	0	0	0	0
B	26.04	0	25	0	0	0	12	19	0	0	0	3	0	0	0
C	73.27	0	0	74	1	0	0	0	0	0	0	0	6	12	1
D	87.23	0	0	2	82	0	0	0	0	0	0	0	4	6	0
E	97.25	1	0	0	0	106	0	0	0	0	0	0	0	0	0
F	64.37	0	2	0	0	0	56	17	1	0	0	0	0	6	0
G	62.89	0	3	0	0	0	15	61	1	0	0	1	0	4	0
H	89.80	0	0	0	1	0	5	2	88	0	0	1	0	1	0
I	76.14	8	0	0	0	0	0	0	0	67	0	0	0	0	0
J	38.00	0	0	0	0	0	0	0	0	0	38	5	0	0	0
K	75.82	0	2	0	0	0	1	2	0	0	2	69	0	0	0
L	78.79	0	0	4	2	0	0	0	0	0	0	0	78	3	0
M	34.04	0	0	18	2	0	5	1	0	0	0	0	4	32	1
N	96.19	0	0	0	0	0	0	0	0	0	0	0	0	0	101
O	50.48	28	0	0	0	0	0	0	0	12	0	0	1	0	0
P	53.92	0	0	7	1	0	3	5	1	0	0	0	4	14	0
Q	73.53	0	0	0	0	0	0	0	0	0	7	1	0	0	0
R	89.90	0	0	1	0	0	0	0	0	0	0	0	2	0	4
S	93.07	0	0	0	2	0	0	0	0	0	0	0	2	0	0
T	85.71	4	0	0	0	0	0	0	0	0	0	0	0	0	0
U	72.97	0	0	0	0	0	1	5	0	0	0	0	15	1	0
V	62.50	0	0	0	0	0	0	3	0	0	0	8	0	0	0
W	45.16	0	25	0	0	0	10	11	1	0	0	3	0	0	0
X	60.95	0	0	0	0	0	0	0	0	0	5	23	0	0	0
Y	74.26	0	0	0	1	0	0	2	0	0	0	2	0	0	0
Z	35.29	0	0	0	0	0	0	0	0	0	34	1	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	68.74	94	57	106	92	107	108	129	92	89	86	117	116	79	107

Table A.8 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	54.08	28	0	0	4	1	1	0	0	0	0	0	0	0	98
B	26.04	0	0	0	0	0	0	0	0	37	0	0	0	0	96
C	73.27	0	5	0	2	0	0	0	0	0	0	0	0	0	101
D	87.23	0	0	0	0	0	0	0	0	0	0	0	0	0	94
E	97.25	0	0	0	0	2	0	0	0	0	0	0	0	0	109
F	64.37	0	2	0	0	0	0	0	0	3	0	0	0	0	87
G	62.89	0	3	0	0	0	0	0	2	7	0	0	0	0	97
H	89.80	0	0	0	0	0	0	0	0	0	0	0	0	0	98
I	76.14	10	0	0	1	2	0	0	0	0	0	0	0	0	88
J	38.00	0	0	15	0	0	0	0	1	0	20	0	21	0	100
K	75.82	0	0	0	0	0	0	0	8	2	5	0	0	0	91
L	78.79	0	2	0	4	0	0	6	0	0	0	0	0	0	99
M	34.04	0	27	0	0	0	0	4	0	0	0	0	0	0	94
N	96.19	0	0	0	4	0	0	0	0	0	0	0	0	0	105
O	50.48	53	0	0	7	2	2	0	0	0	0	0	0	0	105
P	53.92	6	55	0	0	0	0	11	0	0	0	0	0	0	102
Q	73.53	0	0	75	0	0	0	0	2	0	1	0	16	0	102
R	89.90	1	0	0	89	0	2	0	0	0	0	0	0	0	99
S	93.07	0	0	0	0	94	3	0	0	0	0	0	0	0	101
T	85.71	0	0	0	6	3	78	0	0	0	0	0	0	0	91
U	72.97	0	8	0	0	0	0	81	0	0	0	0	0	0	111
V	62.50	0	0	1	0	0	0	0	65	1	0	26	0	0	104
W	45.16	0	0	0	0	0	0	0	1	42	0	0	0	0	93
X	60.95	0	0	4	0	0	0	0	1	6	64	0	2	0	105
Y	74.26	0	0	0	0	0	0	0	21	0	0	75	0	0	101
Z	35.29	0	0	18	0	0	0	0	0	0	13	0	3	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	68.74	92	102	113	117	104	86	102	101	98	103	101	75	101	2674

Table A.9: Classification matrix for validating sample : monoalphabetic 12 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	6' 76	63	[illegible]	0	0	0	0	0	0	11	0	0	0	0	0
B	29.81	0	31	0	0	0	12	11	1	0	0	5	0	0	0
C	62.63	0	0	62	1	0	0	0	1	0	0	0	6	12	0
D	84.91	0	0	4	90	0	0	0	2	0	0	0	2	7	0
E	97.80	2	0	0	0	89	0	0	0	0	0	0	0	0	0
F	61.95	0	4	0	1	0	70	21	3	0	0	0	0	7	3
G	57.28	0	1	0	1	0	25	59	1	0	0	0	0	2	0
H	85.29	0	0	1	5	0	8	0	87	0	0	0	0	1	0
I	65.18	16	0	0	0	0	0	0	0	73	0	0	1	0	0
J	34.00	0	0	0	0	0	0	0	0	0	34	13	0	0	0
K	69.72	0	2	0	0	0	0	1	0	0	1	76	0	0	0
L	75.25	0	0	3	5	0	0	0	0	0	0	0	76	3	0
M	42.45	0	0	13	7	0	2	2	3	0	0	0	2	45	0
N	92.63	0	0	2	0	0	0	0	0	0	0	0	0	0	88
O	53.68	24	0	0	0	0	0	0	0	16	0	0	0	0	0
P	50.00	0	1	3	0	0	10	5	0	0	0	0	4	15	0
Q	67.35	0	0	0	0	0	0	0	0	0	7	3	0	0	0
R	88.12	0	0	1	0	0	0	0	0	1	0	0	2	0	4
S	98.99	0	0	0	0	0	0	0	0	1	0	0	0	0	0
T	88.99	7	0	0	0	0	0	0	0	0	0	0	0	0	0
U	73.03	0	0	0	0	0	[illegible]	6	0	0	0	0	17	1	0
V	65.63	6	2	0	0	0		1	0	0	1	12	0	0	0
W	48.60	0	19	0	0	0	19	13	0	0	0	2	0	0	0
X	67.37	0	0	0	0	0	0	0	0	0	4	19	0	0	0
Y	79.80	0	0	0	0	0	0	3	0	0	0	1	0	0	0
Z	41.84	0	0	0	0	0	0	0	0	0	23	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	68.09	112	60	89	110	89	146	122	98	102	70	131	110	93	92

Table A.9 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	61.76	25	0	0	0	1	2	0	0	0	0	0	0	0	102
B	29.81	0	1	0	0	0	0	0	1	41	0	1	0	0	104
C	62.63	0	12	0	5	0	0	0	0	0	0	0	0	0	99
D	84.91	0	[illegible]	0	0	0	0	0	0	0	0	0	0	0	106
E	97.80	0	0	0	0	0	0	0	0	0	0	0	0	[illegible]	91
F	61.95	0	2	0	0	0	0	0	0	5	0	0	0	0	113
G	57.28	0	3	0	0	0	0	0	1	10	0	0	0	0	103
H	85.29	0	0	0	0	0	0	0	0	0	0	0	0	0	102
I	65.18	19	0	0	1	2	0	0	0	0	0	0	0	0	112
J	34.00	0	0	13	0	0	0	0	0	0	20	0	20	0	100
K	69.72	0	0	3	0	0	0	0	12	6	7	1	0	0	109
L	75.25	0	2	0	2	0	0	10	0	0	0	0	0	0	101
M	42.45	0	29	0	0	0	0	3	0	0	0	0	0	0	106
N	92.63	0	0	0	5	0	0	0	0	0	0	0	0	0	95
O	53.68	51	0	0	2	0	2	0	0	0	0	0	0	0	95
P	50.00	0	49	0	0	0	0	11	0	0	0	0	0	0	98
Q	67.35	0	0	66	0	0	0	0	1	0	1	0	20	0	98
R	88.12	2	0	0	89	0	2	0	0	0	0	0	0	0	101
S	98.99	0	0	0	0	98	0	0	0	0	0	0	0	0	99
T	88.99	0	0	0	3	2	97	0	0	0	0	0	0	0	109
U	73.03	0	0	0	0	0	0	65	0	0	0	0	0	0	89
V	65.63	0	0	0	0	0	0	0	63	0	0	17	0	0	96
W	48.60	0	0	0	0	0	0	0	1	52	1	0	0	0	107
X	67.37	0	0	3	0	0	0	0	1	2	64	0	2	0	95
Y	79.80	0	0	0	0	0	0	0	16	0	0	79	0	0	99
Z	41.84	0	0	18	0	0	0	0	0	1	15	0	41	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	68.09	97	99	103	107	103	103	89	96	117	108	98	83	99	2726

Table A.10: Coefficients of Linear Discrimination for Character Groups, Polyalphabetic Position 1, 1 variable

Group	Coefficients of RELFREQ	CONSTANT
A	1344.39697	-48.84290
B	245.28423	-6.01146
C	699.50891	-15.87003
D	615.84662	-13.13741
E	1920.04333	-96.57281
F	372.67822	7.58959
G	297.68631	-6.40888
H	553.89063	-11.36404
I	1188.30493	-38.86905
J	20.25970	-6.86661
K	99.08076	-5.53144
L	624.53967	-13.37883
M	529.10516	-10.78312
N	1267.06909	-43.79227
O	1286.50378	-45.01196
P	458.63361	-9.13857
Q	33.07404	-6.78558
R	1076.27747	-32.54067
S	1157.65308	-37.06063
T	1464.41003	-57.44825
U	142.59964	-8.76977
V	180.03050	-5.49198
W	232.67303	-5.76444
X	62.08949	-6.02610
Y	244.03333	-5.85013
Z	13.09922	-7.26454
[illegible]	3135.51563	-253.17082

Table A.11: Classification matrix for training sample : polyalphabetic position 1, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.49	24	0	2	0	4	0	0	0	12	0	0	0	0	0
B	0.00	0	0	0	0	0	10	47	0	0	0	0	0	0	0
C	45.54	0	0	46	0	0	0	0	18	0	0	0	21	0	0
D	0.00	0	0	25	0	0	0	0	24	0	0	0	32	0	0
E	86.24	0	0	0	0	94	0	0	0	0	0	0	0	0	0
F	9.20	0	0	0	0	0	8	19	10	0	0	0	5	1	0
G	43.30	0	0	0	0	0	10	42	4	0	0	0	0	0	0
H	27.55	0	0	26	0	0	1	2	27	0	0	0	22	0	0
I	7.95	15	0	3	0	2	0	0	0	7	0	0	0	0	1
J	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	0	3	0	0	0	0	0	0	0
L	30.30	0	0	32	0	0	1	1	27	0	0	0	30	0	0
M	2.13	0	0	14	0	0	0	4	25	0	0	0	19	2	0
N	0.00	28	0	0	0	0	0	0	0	14	0	0	0	0	0
O	16.19	26	0	1	0	4	0	0	0	10	0	0	0	0	3
P	1.96	0	0	5	0	0	3	7	30	0	0	0	13	2	0
Q	0.00	6	0	0	0	0	0	0	0	0	0	0	0	0	0
R	45.45	2	0	9	0	0	0	0	0	15	0	0	0	0	1
S	12.87	7	0	8	0	1	0	0	0	22	0	0	0	0	1
T	38.46	25	0	1	0	17	0	0	0	6	0	0	0	0	0
U	40.54	0	0	2	0	0	2	10	37	0	0	0	7	3	0
V	14.42	0	0	0	0	0	3	30	0	0	0	0	0	0	0
W	7.53	0	0	0	0	0	8	35	1	0	0	0	0	0	0
X	0.00	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Y	22.77	0	0	0	0	0	5	53	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	21.80	117	0	174	0	122	51	256	203	86	0	0	149	8	6

Table A.11 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	24.49	15	0	0	6	4	31	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	12	5	4	0	18	0	0	96
C	45.54	0	0	0	13	1	0	2	0	0	0	0	0	0	101
D	0.00	0	0	0	6	0	0	7	0	0	0	0	0	0	94
E	86.24	0	0	0	0	0	13	0	0	0	0	0	0	2	109
F	9.20	0	1	0	0	0	0	40	0	1	0	2	0	0	87
G	43.30	0	1	0	0	0	0	27	0	2	0	11	0	0	97
H	27.55	0	0	0	0	0	0	18	0	0	0	2	0	0	98
I	7.95	9	0	0	26	14	11	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	99	1	0	0	0	0	100
K	0.00	0	1	0	0	0	0	0	52	22	0	13	0	0	91
L	30.30	0	1	0	3	1	0	3	0	0	0	0	0	0	99
M	2.13	0	2	0	3	0	0	25	0	0	0	0	0	0	94
N	0.00	15	0	0	14	15	19	0	0	0	0	0	0	0	105
O	16.19	17	0	0	15	8	21	0	0	0	0	0	0	0	105
P	1.96	0	2	0	0	0	0	40	0	0	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	95	4	0	3	0	0	102
R	45.45	12	0	0	45	14	1	0	0	0	0	0	0	0	99
S	12.87	13	0	0	27	13	9	0	0	0	0	0	0	0	101
T	38.46	11	0	0	4	2	35	0	0	0	0	0	0	0	91
U	40.54	0	5	0	0	0	0	45	0	0	0	0	0	0	111
V	14.42	0	0	0	0	0	0	5	15	23	0	28	0	0	104
W	7.53	0	0	0	0	0	0	10	5	7	0	27	0	0	93
X	0.00	0	0	0	0	0	0	0	85	14	0	3	0	0	105
Y	22.77	0	0	0	0	0	0	13	2	5	0	23	0	0	101
Z	0.00	0	0	0	0	0	0	0	102	0	0	0	0	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	21.80	92	13	0	162	72	140	247	460	83	0	130	0	103	2674

Table A.12: Classification matrix for validating sample : polyalphabetic position 1, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	20.59	21	0	0	0	4	0	0	0	10	0	0	0	0	0
B	0.00	0	0	0	0	0	6	47	1	0	0	0	1	0	0
C	44.44	0	0	44	0	0	0	1	13	0	0	0	20	1	0
D	0.00	0	0	27	0	0	0	2	29	0	0	0	32	0	0
E	85.71	1	0	0	0	78	0	0	0	0	0	0	0	0	0
F	6.19	0	0	1	0	0	7	34	9	0	0	0	4	0	0
G	40.78	0	0	0	0	0	7	42	3	0	0	0	1	1	0
H	15.69	0	0	27	0	0	1	2	16	0	0	0	31	1	0
I	18.75	19	0	4	0	2	0	0	0	21	0	0	0	0	0
J	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	0	10	0	0	0	0	0	0	0
L	30.69	0	0	40	0	0	2	0	19	0	0	0	31	0	0
M	0.00	0	0	14	0	0	2	2	36	0	0	0	23	0	0
N	1.05	16	0	2	0	3	0	0	0	15	0	0	0	0	1
O	17.89	16	0	0	0	1	0	0	0	16	0	0	0	0	1
P	3.06	0	0	4	0	0	7	8	26	0	0	0	12	1	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	33.66	10	0	13	0	1	0	0	0	17	0	0	0	0	1
S	14.14	22	0	3	0	0	0	0	0	22	0	0	0	0	1
T	56.88	16	0	0	0	13	0	0	0	5	0	0	0	0	0
U	31.46	0	0	4	0	0	1	13	35	0	0	0	8	0	0
V	15.63	0	0	0	0	0	1	28	0	0	0	0	0	0	0
W	11.21	0	0	0	0	0	8	40	0	0	0	0	0	0	0
X	0.00	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Y	21.21	0	0	0	0	0	6	41	1	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	1	0	0	0	0	0	0	0
@	98.99	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	20.73	121	0	183	0	103	48	272	188	106	0	0	163	4	4

Table A.12 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	20.59	9	0	0	5	6	47	0	0	0	0	0	0	0	102
B	0.00	0	1	0	0	0	0	6	2	12	0	28	0	0	10[illegible]
C	44.44	0	0	0	10	0	0	10	0	0	0	0	0	0	99
D	0.00	0	1	0	5	2	0	8	0	0	0	0	0	0	106
E	85.71	0	0	0	0	0	12	0	0	0	0	0	0	0	91
F	6.19	0	3	0	0	0	0	48	2	1	0	4	0	0	113
G	40.78	0	1	0	0	0	0	32	1	3	0	12	0	0	103
H	15.69	0	[illegible]	0	2	0	0	21	0	0	0	0	0	0	102
I	18.75	16	[illegible]	0	20	17	13	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	100	0	0	0	0	0	100
K	0.00	0	0	0	0	0	0	0	58	22	0	19	0	0	109
L	30.69	0	1	0	3	0	0	5	0	0	0	0	0	0	101
M	0.00	0	1	0	0	0	0	28	0	0	0	0	0	0	106
N	1.05	12	0	0	17	15	14	0	0	0	0	0	0	0	95
O	17.89	17	0	0	16	9	19	0	0	0	0	0	0	0	95
P	3.06	0	3	0	1	0	0	34	0	1	0	1	0	0	98
Q	0.00	0	0	0	0	0	0	0	93	3	0	2	0	0	98
R	33.66	7	0	0	34	13	5	0	0	0	0	0	0	0	101
S	14.14	12	0	0	16	14	9	0	0	0	0	0	0	0	99
T	56.88	7	0	0	3	3	62	0	0	0	0	0	0	0	109
U	[illegible]1.46	0	0	0	0	0	0	28	0	0	0	0	0	0	89
V	15.63	0	0	0	0	0	0	0	15	25	0	27	0	0	96
W	11.21	0	0	0	0	0	0	15	5	12	0	27	0	0	107
X	0.00	0	0	0	0	0	0	0	78	15	0	1	0	0	95
Y	21.21	0	0	0	0	0	0	18	3	9	0	21	0	0	99
Z	0.00	0	0	0	0	0	0	0	95	2	0	0	0	0	98
@	98.99	0	0	0	0	0	0	0	0	0	0	0	0	98	99
TOTAL	20.73	80	12	0	132	79	181	253	452	105	0	142	0	98	2726

Table A.13: Coefficients of Linear Discriminant Functions for Character Groups : Polyalphabetic, Position 1, 3 Variables

No.	Group	RELFREQ	INFOCONT	CTOENTRO	CONSTANT
1	A	-1918.92542	2.24810	1719.08252	-97.27392
2	B	-1589.57288	3.01248	983.69550	-27.87043
3	C	-2459.15771	2.29879	1665.15271	-61.78775
4	D	-2414.97852	2.35522	1599.22156	-56.07608
5	E	-489.63416	2.54624	1278.06543	-126.46337
6	F	-2007.72729	2.68380	1264.18909	-37.53569
7	G	-1790.94739	2.84861	1114.06726	-31.68476
8	H	-2331.93896	2.42683	1524.51807	-51.11260
9	I	-2165.64380	2.21428	1765.88135	-89.59039
10	J	-117.19640	2.98739	100.72247	-13.35888
11	K	-761.81506	3.85440	485.43054	-20.20765
12	L	-2419.60083	2.34970	1606.09216	-56.62347
13	M	-2299.31323	2.45243	1494.91223	-49.30126
14	N	-2069.40039	2.22121	1756.85925	-94.06876
15	O	-2019.72815	2.23218	1741.24170	-94.51987
16	P	-2203.95728	2.53266	1409.45935	-44.27585
17	Q	-225.76466	3.38054	167.69498	-15.41019
18	R	-2332.51855	2.19365	1794.20264	-84.67841
19	S	-2210.78906	2.2089[illegible]	1773.36609	-88.15432
20	T	-1658.11487	2.29835	1646.35315	-102.43700
21	U	-2180.09888	2.55235	1388.90613	-43.12753
22	V	-1275.93152	3.28864	789.35529	-23.36235
23	W	-1530.25549	3.06224	946.77625	-26.76438
24	X	-482.72217	3.52565	317.83447	-16.68410
25	Y	-1592.82520	3.00609	984.67377	-27.70877
26	Z	-61.60412	2.42714	62.60467	-11.48625
27	@	-4336.22314	3.70690	-588.14282	-266.64636

Table A.14: Classification matrix for training sample : polyalphabetic position 1, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	23.47	23	0	2	0	4	0	0	0	12	0	0	0	0	3
B	0.00	0	0	0	0	0	8	31	0	0	0	6	0	0	0
C	44.55	0	0	45	0	0	0	0	16	0	0	0	19	2	0
D	3.19	0	0	26	3	0	1	0	21	0	0	0	28	3	0
E	82.57	0	0	0	0	90	0	0	0	0	0	0	0	0	0
F	19.54	0	0	0	1	0	17	14	4	0	0	1	4	4	0
G	38.14	0	0	0	0	0	16	37	1	0	0	0	0	3	0
H	18.37	0	0	27	0	0	3	3	18	0	0	0	20	10	0
I	7.95	19	0	3	0	2	0	0	0	7	0	0	0	0	5
J	0.00	0	0	0	0	0	0	0	0	0	0	20	0	0	0
K	53.85	0	0	0	0	0	0	1	0	0	0	49	0	0	0
L	29.29	0	0	32	0	0	1	1	24	0	0	0	29	3	0
M	7.45	0	0	15	2	0	4	3	17	0	0	0	15	7	0
N	2.86	29	0	0	0	0	0	0	0	14	0	0	0	0	3
O	18.10	24	0	1	0	3	0	0	0	11	0	0	0	0	3
P	8.82	0	0	6	1	0	12	9	18	0	0	0	11	9	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	28	0	0	0
R	47.47	1	0	9	0	0	0	0	0	15	0	0	0	0	3
S	13.86	7	0	8	0	0	0	0	0	22	0	0	0	0	4
T	38.46	15	0	1	0	17	0	0	0	6	0	0	0	0	1
U	27.03	0	0	4	1	0	13	8	23	0	0	0	4	13	0
V	40.38	0	0	0	0	0	2	10	0	0	0	15	0	0	0
W	19.35	0	0	0	0	0	11	22	1	0	0	6	0	0	0
X	27.62	0	0	0	0	0	0	0	0	0	0	47	0	0	0
Y	19.80	0	0	0	0	0	10	29	0	0	0	3	0	0	0
Z	50.00	0	0	0	0	0	0	0	0	0	0	12	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	27.79	118	0	179	8	116	98	168	143	87	0	187	130	54	22

Table A.14 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	23.47	14	0	0	7	3	30	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	6	13	15	0	17	0	0	96
C	44.55	0	0	0	17	0	0	2	0	0	0	0	0	0	101
D	3.19	0	2	0	6	0	0	4	0	0	0	0	0	0	94
E	82.57	0	0	0	0	0	17	0	0	0	0	0	0	2	109
F	19.54	0	6	0	0	0	0	25	0	3	0	8	0	0	87
G	38.14	0	1	0	0	0	0	13	6	12	0	8	0	0	97
H	18.[illegible]	0	2	0	0	0	0	13	0	2	0	0	0	0	98
I	7.95	5	0	0	27	13	7	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	1	0	42	0	37	0	100
K	53.85	0	1	0	0	0	0	0	21	5	13	1	0	0	91
L	29.29	0	2	0	4	1	0	2	0	0	0	0	0	0	99
M	7.45	0	7	0	3	0	0	20	0	1	0	0	0	0	94
N	2.86	13	0	0	15	14	17	0	0	0	0	0	0	0	105
O	18.10	19	0	0	15	9	20	0	0	0	0	0	0	0	105
P	8.82	0	9	0	0	0	0	26	0	0	0	1	0	0	102
Q	0.00	0	0	0	0	0	0	0	6	0	42	0	26	0	102
R	47.47	11	0	0	47	12	1	0	0	0	0	0	0	0	99
S	13.86	10	0	0	27	14	9	0	0	0	0	0	0	0	101
T	38.46	10	0	0	4	2	35	0	0	0	0	0	0	0	91
U	27.03	0	12	0	0	0	0	30	0	1	0	2	0	0	111
V	40.38	0	1	0	0	0	0	2	42	15	1	16	0	0	104
W	19.35	0	0	0	0	0	0	2	18	18	0	15	0	0	93
X	27.62	0	0	0	0	0	0	0	10	1	29	3	15	0	105
Y	19.80	0	0	0	0	0	0	6	16	17	0	20	0	0	101
Z	50.00	0	0	0	0	0	0	0	0	0	39	0	51	0	102
@	100.00	0	0	0	0	9	0	0	0	0	0	0	0	101	101
TOTAL	27.79	82	43	0	172	68	136	151	133	90	166	91	129	103	2674

Table A.15: Classification matrix for validating sample : polyalphabetic position 1, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	22.55	23	0	0	0	4	0	0	0	9	0	0	0	0	1
B	0.00	0	0	0	0	0	4	26	0	0	0	4	1	1	0
C	45.45	0	0	45	0	0	2	1	9	0	0	0	19	3	0
D	0.94	0	0	30	1	0	1	2	20	0	0	0	28	6	0
E	82.42	1	0	0	0	75	0	0	0	0	0	0	0	0	0
F	23.01	0	0	1	0	0	26	28	6	0	0	3	4	1	0
G	29.13	0	0	0	0	0	20	30	2	0	0	1	1	1	0
H	9.80	0	0	28	3	0	4	2	10	0	0	0	27	4	0
I	18.75	18	0	4	0	2	0	0	0	21	0	0	0	0	4
J	0.00	0	0	0	0	0	0	0	0	0	0	14	0	0	0
K	39.45	0	0	0	0	0	0	1	0	0	0	43	0	0	0
L	29.70	0	0	39	0	0	3	0	15	0	0	0	30	3	0
M	3.77	0	0	15	2	0	7	3	29	0	0	0	20	4	0
N	3.16	15	0	2	0	3	0	0	0	14	0	0	0	0	3
O	16.84	17	0	0	0	1	0	0	0	16	0	0	0	0	2
P	6.12	0	0	4	6	0	16	7	21	0	0	1	12	5	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	18	0	0	0
R	33.66	10	0	13	0	0	0	0	0	15	0	0	0	0	2
S	14.14	20	0	3	0	0	0	0	0	22	0	0	0	0	3
T	54.13	20	0	0	0	12	0	0	0	5	0	0	0	0	1
U	17.98	0	0	4	1	0	9	10	24	0	0	0	7	10	0
V	35.42	0	0	0	0	0	0	9	0	0	0	21	0	0	0
W	16.82	0	0	0	0	0	11	24	0	0	0	7	0	0	0
X	28.42	0	0	0	0	0	0	1	0	0	0	52	0	0	0
Y	16.16	0	0	0	0	0	13	26	0	0	0	5	0	0	0
Z	48.98	0	0	0	0	0	8	1	0	0	0	9	0	0	0
@	98.99	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	25.57	124	0	188	7	98	116	171	136	102	0	178	149	38	16

Table A.15 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	22.55	8	0	0	5	7	45	0	0	0	0	0	0	0	102
B	0.00	0	1	0	0	0	0	5	21	26	0	15	0	0	104
C	45.45	0	2	0	0	0	0	8	0	0	0	0	0	0	99
D	0.94	0	5	0	6	1	0	6	0	0	0	0	0	0	106
E	82.42	0	0	0	0	0	15	0	0	0	0	0	0	0	91
F	23.01	0	7	0	0	0	0	26	0	7	0	4	0	0	113
G	29.13	0	2	0	0	0	0	18	11	6	0	11	0	0	103
H	9.80	0	7	0	2	0	0	15	0	0	0	0	0	0	102
I	18.75	13	0	0	22	15	13	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	0	0	54	0	32	0	100
K	39.45	0	0	0	0	0	0	0	31	7	21	4	2	0	109
L	29.70	0	4	0	5	0	0	2	0	0	0	0	0	0	101
M	3.77	0	5	0	0	0	0	21	0	0	0	0	0	0	106
N	3.16	12	0	0	18	15	13	0	0	0	0	[illegible]	0	0	95
O	16.84	16	0	0	16	9	18	0	0	0	0	[illegible]	0	0	95
P	6.12	0	6	0	1	0	0	22	0	1	0	2	0	0	98
Q	0.00	0	0	0	0	0	0	0	3	0	50	0	27	0	98
R	33.66	6	0	0	34	16	5	0	0	0	0	0	0	0	101
S	14.14	11	0	0	17	14	9	0	0	0	0	0	0	0	99
T	54.13	6	0	0	3	3	59	0	0	0	0	0	0	0	109
U	17.98	0	5	0	0	0	0	16	0	0	0	3	0	0	89
V	35.42	0	0	0	0	0	0	0	34	17	0	14	1	0	96
W	16.82	0	0	0	0	0	0	7	24	18	0	16	0	0	107
X	28.42	0	0	0	0	0	0	0	7	0	27	0	8	0	95
Y	16.16	0	1	0	0	0	0	8	18	12	0	16	0	0	99
Z	48.98	0	0	0	0	0	0	0	1	0	39	0	48	0	98
@	98.99	0	0	0	0	0	0	0	0	0	0	0	0	98	99
TOTAL	25.57	72	45	0	139	80	177	154	150	94	191	85	118	98	2726

Table A.16: Coeficients of linear discriminant functions for character groups : polyalphabetic, position 1, 20 variables

No.	Variable	A	B	C
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	-35353.67188	258724.40625	238113.06250
3	WINROWS3	-5378.59131	-22180.05859	-27146.04102
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	2.48928	1.15429	13.11606
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	2.04619	0.49993	1.17099
8	CTOENTRO	1534.00293	1867.39856	2523.90912
9	RCTOJEN3	-599.43890	-297.02103	293.82620
10	CCTOJEN3	8.72361	-156.28743	69.27608
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	-0.07631	8.08837	2.11441
13	WINROWSS	-10981.17090	-8660.66699	-19514.37305
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-0.70233	-403.52948	-447.30484
16	RWCENTRO	11.59558	20.05525	18.45724
17	WINCOLSS	31.63961	-10210.90332	-15387.84863
18	CWCPROSS	3.44627	7.79503	-10.22793
19	CCTOJENT	321.57742	-154.31703	-299.66312
20	CWCENTRO	22.21303	9.36079	16.93430
	CONSTANT	-131.53845	-49.57996	81.50282

Table A.16 (continued)

No.	Variable	D	E	F
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	269186.25000	95546.57813	289618.15625
3	WINROWS3	-20094.77539	-77864.37500	23161.46289
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	9.17036	-5.58558	8.80787
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	1.94010	2.36759	0.71627
8	CTOENTRO	2650.82178	1234.89734	2284.12891
9	RCTOJEN3	-150.15184	-860.39807	-268.68845
10	CCTOJEN3	-130.36780	-74.39330	-126.77068
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	3.35129	0.77921	7.38933
13	WINROWSS	-27535.69531	59143.91797	14694.88477
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-738.14185	583.98029	451.62738
16	RWCENTRO	12.92083	6.34[illegible]13	18.22694
17	WINCOLSS	-15725.59277	-44987.88281	16668.37891
18	CWCPROSS	-11.82721	2.99093	0.78222
19	CCTOJENT	-194.73984	229.24832	318.01193
20	CWCENTRO	17.65148	23.52200	11.93639
	CONSTANT	-78.94025	187.52400	59.97687

Table A.16 (continued)

No.	Variable	G	H	I
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	264535.53125	276304.78125	-51673.57813
3	WINROWS3	-21538.38867	-10639.28223	11012.80078
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	3.76666	5.44269	8.02890
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	1.69706	0.74409	0.98008
8	CTOENTRO	2070.12207	2505.04688	1563.92969
9	RCTOJEN3	-267.27444	-52.15048	-766.93481
10	CCTOJEN3	-194.02052	-10.02482	74.03851
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	9.04081	7.81077	1.30297
13	WINROWSS	-16048.72949	-29080.12500	-8842.55176
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-449.57745	-708.93439	-149.93600
16	RWCENTRO	13.43523	12.22487	14.41866
17	WINCOLSS	-9978.59375	-16782.16359	-1601.44568
18	CWCPROSS	-2.11559	1.51044	-1.36750
19	CCTOJENT	-156.45291	-437.36826	507.32242
20	CWCENTRO	13.17125	11.75691	24.05477
	CONSTANT	-54.[illegible]58	-69.71219	-137.06837

Table A.16 (continued)

No	Variable	J	K	L
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	18468.65820	150568.78125	220290.70313
3	WINROWS3	-1147.82288	-12995.47070	-25664.28711
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	6.96981	-11.96440	16.[illegible]5530
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	2.16410	5.66868	0.77250
8	CTOENTRO	165.25398	1140.54846	2374.84424
9	RCTOJEN3	4.02677	-214.10696	-278.08710
10	CCTOJEN3	-28.21992	-9.86453	-191.69933
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	1.21468	-1.56454	2.31574
13	WINROWSS	-1173.32825	-8519.76172	-25102.62305
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-37.30650	-267.61703	-592.79712
16	RWCENTRO	1.16839	8.04270	20.44022
17	WINCOLSS	-615.47949	-3687.10181	-7811.13428
18	CWCPROSS	-0.78732	-6.49305	-10.08385
19	CCTOJENT	-15.15695	-97.87482	12.07679
20	CWCENTRO	1.30826	10.51879	20.64241
	CONSTANT	-13.55187	-30.28959	-87.11555

Table A.16 (continued)

No	Variable	M	N	O
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	268396.68750	135002.73438	16209.26563
3	WINROWS3	-28193.66211	-12771.21680	37701.03125
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	14.77130	16.22761	5.40650
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	1.05661	3.74446	2.11108
8	CTOENTRO	2530.31543	2093.68799	1721.98132
9	RCTOJEN3	-345.25424	154.70464	-512.83069
10	CCTOJEN3	-156.68546	-13.29615	-146.26846
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	2.96889	-0.74462	0.90220
13	WINROWSS	-21088.36914	451.47757	-76.43085
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-528.93402	218.05121	128.28659
16	RWCENTRO	19.90871	10.93983	10.32422
17	WINCOLSS	-10958.27734	-34907.44922	15179.72363
18	CWCPROSS	-10.46547	-29.63243	7.37757
19	CCTOJENT	-160.87224	1043.14160	157.23146
20	CWCENTRO	16.10753	15.26870	21.90661
	CONSTANT	-74.60822	-120.46352	127.86937

Table A.16 (continued)

No.	Variable	P	Q	R
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	260541.37500	19149.39844	129057.10938
3	WINROWS3	-26310.51953	-6109.27295	26423.72070
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	15.14336	-4.13231	12.81312
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	0.76388	4.50582	2.17210
8	CTOENTRO	2347.06201	271.21848	2140.93613
9	RCTOJEN3	-300.27924	-85.54340	180.83118
10	CCTOJEN3	-180.93654	-101.78305	152.29263
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	3.28175	4.56068	0.63458
13	WINROWSS	-17288.90234	1646.91431	-3730.74121
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-507.14415	158.63974	39.60451
16	RWCENTRO	22.57870	8.08010	13.40697
17	WINCOLSS	-10735.03027	2295.37988	23552.11719
18	CWCPROSS	-8.09376	5.24886	15.81204
19	CCTOJENT	-149.91797	56.79264	348.52304
20	CWCENTRO	16.97687	1.75946	19.06349
	CONSTANT	72.20391	-16.50293	108.26702

Table A.16 (continued)

No.	Variable	S	T	U
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	107548.83594	111952.67969	250635.85938
3	WINROWS3	-41493.00781	-10669.08301	-40688.83594
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	8 36564	6.23935	20.85389
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	1.79858	3.43894	0.17638
8	CTOENTRO	2057.26416	2027.29187	2331.70435
9	RCTOJEN3	-314.63785	-252.98839	-573.83813
10	CCTOJEN3	-67.37389	378.05231	-164.25404
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.35743	-6.56690	3.49624
13	WINROWSS	17715.46289	-16747.71680	-13656.02344
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-413.28073	-464.16263	-486.68100
16	RWCENTRO	13.02247	15.67281	24.34744
17	WINCOLSS	-21051.66602	-16112.32715	-419.44510
18	CWCPROSS	-8.51061	-17.50638	-9.56832
19	CCTOJENT	112.96993	-268.05231	93.81648
20	CWCENTRO	22.10063	19.53738	21.52815
	CONSTANT	-120.65246	-126.26608	-81.71071

Table A.16 (continued)

No.	Variable	V	W	X
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	212071.73438	251456.17188	88069 56250
3	WINROWS3	-21468.56836	-19729.27148	-7600.45361
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	-12.85909	-1.95207	-17.85978
6	CWCPROS3	0.00000	0 00000	0.00000
7	INFOCONT	6.21964	0.91888	3.68650
8	CTOENTRO	1733 96216	1816.01941	655.87738
9	RCTOJEN3	-306.06113	-265.03049	-155.44960
10	CCTOJEN3	-188 50824	-160.61069	-5.49261
11	RWCENTR3	0 00000	0.00000	0.00000
12	CWCENTR3	5.95294	7.79470	-0.65636
13	WINROWSS	-15444.11719	-9683.75977	-5343.41406
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-245.42442	-382.55737	-219.25607
16	RWCENTRO	-2.51156	17.26003	8.42920
17	WINCOLSS	-5169 95752	-11152.43652	-886.03888
18	CWCPROSS	-15.72805	11.14947	13.05217
19	CCTOJENT	-108.71886	-168.72932	35.93974
20	CWCENTRO	15.05891	8.99617	3.20655
	CONSTANT	-42.70847	-46.11631	-20.84834

Table A.16 (continued)

No.	Variable	Y	Z	ω
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	234544.26563	7724.83154	1074312.25000
3	WINROWS3	-26263.70117	-249.63757	-102838.23438
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	-3.93827	0.44413	13.22870
6	CWCPROS3	0.00000	0.00000	0.00000
7	INFOCONT	4.99424	2.16647	2.51588
8	CTOENTRO	1994.45081	98.79552	2931.88232
9	RCTOJEN3	-353.13547	-0.35160	-617.42810
10	CCTOJEN3	-290.68857	28.21865	-1006.75238
11	RWCENTR3	0.00000	0.00000	0.[illegible]000
12	CWCENTR3	12.88672	-1.52487	18.[illegible]8070
13	WINROWSS	20830.97656	-2351.89233	25977.82422
14	RWCPROSS	0.00000	0.00000	0.00000
15	RCTOJENT	-302.09720	-66.42678	1874.83887
16	RWCENTRO	-7.43027	1.82251	-16.13121
17	WINCOLSS	-2311.71924	-163.86378	-126071.61719
18	CWCPROSS	-17.94719	1.57592	-21.47664
19	CCTOJENT	-19.00233	3.05641	-1420.41882
20	CWCENTRO	18.41440	0.45812	23.64644
	CONSTANT	-55.72183	-11.55112	433.74457

Table A.17: Classification matrix for training sample : polyalphabetic position 1, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	57.14	56	0	1	0	0	0	0	0	11	0	0	1	0	0
B	34.38	0	33	0	0	0	15	13	0	0	0	5	0	1	0
C	55.45	0	0	56	6	0	0	0	2	0	0	0	15	9	1
D	79.79	0	0	6	75	0	0	1	0	0	0	0	5	7	0
E	95.41	1	0	0	0	104	0	0	0	1	0	0	0	0	0
F	37.93	0	15	3	3	0	33	11	10	0	0	0	0	3	0
G	51.55	0	5	0	1	0	9	50	4	0	0	1	0	5	0
H	82.65	0	1	1	4	0	7	2	81	0	0	0	0	0	0
I	78.41	11	0	0	0	0	0	0	0	69	0	0	3	0	0
J	36.00	0	0	0	0	0	0	0	0	0	36	5	0	0	0
K	45.05	0	1	0	0	0	0	8	0	0	10	41	0	1	0
L	66.67	0	0	6	2	0	0	0	0	0	0	0	66	7	0
M	25.53	0	0	18	4	0	6	2	0	0	0	0	8	24	0
N	93.33	0	0	0	0	0	0	0	0	0	0	0	0	0	98
O	77.14	4	0	0	0	1	0	0	0	0	0	0	1	0	0
P	46.08	0	0	8	5	0	7	3	0	0	0	0	12	11	0
Q	31.37	0	0	0	0	0	0	0	0	0	29	0	0	0	0
R	81.82	0	0	4	0	0	0	0	0	0	0	0	2	0	4
S	83.17	0	0	0	5	2	0	0	0	0	0	0	1	0	1
T	79.12	1	0	1	0	0	0	0	0	2	0	0	0	0	2
U	72.97	0	0	0	0	0	1	1	0	0	0	0	12	6	0
V	51.92	0	0	0	3	0	0	6	0	0	2	11	0	0	0
W	36.56	0	28	0	0	0	11	10	0	0	1	3	0	1	0
X	38.10	0	0	0	0	0	0	0	0	0	14	21	0	0	0
Y	75.25	0	0	0	1	0	0	5	0	0	0	1	0	2	0
Z	50.00	0	0	0	0	0	0	0	0	0	32	2	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	61.93	73	83	104	109	107	89	112	97	83	124	90	126	77	106

Table A.17 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	57.14	22	9	0	2	2	3	0	0	0	0	0	0	0	98
B	34.38	0	1	0	0	0	0	1	2	25	0	0	0	0	96
C	55.45	0	6	0	6	0	0	0	0	0	0	0	0	0	101
D	79.79	0	0	0	0	0	0	0	0	0	0	0	0	0	94
E	95.41	2	0	0	0	1	0	0	0	0	0	0	0	0	109
F	37.93	0	6	0	0	0	0	0	1	2	0	0	0	0	87
G	51.55	0	10	0	0	0	0	0	5	5	0	2	0	0	97
H	82.65	0	0	0	0	0	0	0	0	1	0	1	0	0	98
I	78.41	4	0	0	0	1	0	0	0	0	0	0	0	0	88
J	36.00	0	0	9	0	0	0	0	0	0	13	0	37	0	100
K	45.05	0	0	1	0	0	0	0	14	3	11	1	0	0	91
L	66.67	0	8	0	5	0	0	5	0	0	0	0	0	0	99
M	25.53	0	21	0	1	0	0	9	0	0	0	1	0	0	94
N	93.33	0	0	0	5	0	2	0	0	0	0	0	0	0	105
O	77.14	81	0	0	13	5	0	0	0	0	0	0	0	0	105
P	46.08	0	47	0	0	0	0	9	0	0	0	0	0	0	102
Q	31.37	0	0	32	0	0	0	0	3	1	10	1	26	0	102
R	81.82	2	0	0	81	0	6	0	0	0	0	0	0	0	99
S	83.17	1	0	0	6	84	1	0	0	0	0	0	0	0	101
T	79.12	5	0	0	7	1	72	0	0	0	0	0	0	0	91
U	72.97	0	9	0	0	0	0	81	0	0	0	1	0	0	111
V	51.92	0	0	0	0	0	0	0	54	0	2	26	0	0	104
W	36.56	0	1	0	0	0	0	0	3	34	1	0	0	0	93
X	38.10	0	0	3	0	0	0	0	1	11	40	0	15	0	105
Y	75.25	0	0	0	0	0	0	0	16	0	0	76	0	0	101
Z	50.00	0	0	2	0	0	0	0	0	0	15	0	51	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	61.93	117	109	47	126	94	84	105	99	82	92	109	129	101	2674

Table A.18: Classification matrix for validating sample : polyalphabetic position 1, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	65.69	67	0	0	0	0	0	0	0	13	0	0	0	0	0
B	40.38	0	42	0	0	0	10	5	1	0	0	4	0	1	0
C	52.53	0	0	52	6	0	0	1	2	0	0	0	6	9	0
D	77.36	0	0	5	82	0	2	1	3	0	0	0	1	8	0
E	96.70	0	0	0	0	88	0	0	0	0	0	0	[illegible]	0	0
F	34.51	0	12	2	2	0	39	26	6	0	2	1	6	6	0
G	44.66	0	8	0	2	0	14	46	4	0	0	3	0	3	0
H	80.39	0	0	2	5	0	4	1	82	0	0	0	0	4	0
I	72.32	23	0	0	0	0	0	0	0	81	0	0	3	0	0
J	38.00	0	0	0	0	0	0	0	0	0	38	7	0	0	0
K	39.45	0	7	0	0	0	0	8	0	0	15	43	0	0	0
L	71.29	0	0	6	6	0	0	0	0	0	0	0	72	4	0
M	27.36	0	1	21	5	0	5	1	1	0	0	0	12	29	0
N	87.37	0	0	2	0	0	0	0	0	0	0	0	0	0	83
O	73.68	10	0	0	0	0	0	0	0	2	0	0	0	0	0
P	38.78	0	2	5	1	0	10	4	1	0	0	0	10	18	0
Q	26.53	0	0	0	0	0	0	0	0	0	36	2	0	0	0
R	68.32	0	0	3	2	0	0	0	0	0	0	0	3	0	10
S	93.94	0	0	0	3	0	0	0	0	0	0	0	0	0	0
T	77.06	1	0	0	1	0	0	0	0	2	0	0	0	0	1
U	71.91	0	0	0	1	0	0	1	0	0	0	0	13	3	0
V	44.79	0	2	0	0	0	0	4	0	0	0	11	0	0	0
W	27.10	0	29	0	0	0	20	12	0	0	1	9	0	1	0
X	45.26	0	1	0	0	0	0	0	0	0	19	16	0	0	0
Y	70.71	0	0	0	3	0	0	5	0	0	0	2	0	1	0
Z	48.98	0	0	0	0	0	0	0	0	0	34	3	0	0	0
ω	98.99	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	59.39	101	104	98	119	89	104	115	100	98	145	101	120	87	94

Table A.18 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	65.69	19	0	0	2	0	1	0	0	0	0	0	0	0	102
B	40.38	0	3	0	0	0	0	1	4	31	2	0	0	0	104
C	52.53	0	13	0	9	0	0	1	0	0	0	0	0	0	99
D	77.36	0	1	0	0	1	0	1	0	0	0	1	0	0	106
E	96.70	3	0	0	0	0	0	0	0	0	0	0	0	0	91
F	34.51	0	13	0	0	0	0	0	0	4	0	0	0	0	113
G	44.66	0	10	0	0	0	0	1	1	11	0	0	0	0	103
H	80.39	0	1	0	0	0	0	0	0	3	0	0	0	0	102
I	72.32	3	0	0	0	1	1	0	0	0	0	0	0	0	112
J	38.00	0	0	10	0	0	0	0	0	0	13	0	12	0	100
K	39.45	0	0	4	0	0	0	0	16	6	6	2	2	0	109
L	71.29	0	5	0	1	0	1	6	0	0	0	0	0	0	101
M	27.36	0	26	0	0	0	0	5	0	0	0	0	0	0	106
N	87.37	0	0	0	9	0	1	0	0	0	0	0	0	0	95
O	73.68	70	0	0	13	0	0	0	0	0	0	0	0	0	95
P	38.78	0	38	0	0	0	0	8	1	0	0	0	0	0	98
Q	26.53	0	0	26	0	0	0	0	1	0	6	0	27	0	98
R	68.32	9	0	0	69	2	3	0	0	0	0	0	0	0	101
S	93.94	0	0	0	2	93	1	0	0	0	0	0	0	0	99
T	77.06	6	0	0	11	3	84	0	0	0	0	0	0	0	109
U	71.91	0	6	0	0	0	0	64	9	0	0	1	0	0	89
V	44.79	0	0	1	0	0	0	0	43	1	1	32	1	0	96
W	27.10	0	1	0	0	0	0	0	4	29	1	0	0	0	107
X	45.26	0	0	4	0	0	0	0	0	4	43	0	8	0	95
Y	70.71	0	0	0	0	0	0	0	17	1	0	70	0	0	99
Z	48.98	0	0	4	0	0	0	0	0	1	8	0	48	0	98
@	98.99	0	0	0	0	0	0	0	0	0	0	0	0	98	99
TOTAL	59.39	110	117	49	116	100	92	87	87	91	80	106	118	98	2726

Table A.19: Coefficients of linear discrimination for character groups, polyalphabetic, position 2, 1 variable

Group	Coefficients of RELFREQ	CONSTANT
A	1272.93518	-47.56224
B	217.00815	-5.77904
C	633.62518	-14.49211
D	568.60364	-12.40545
E	1772.29565	-89.41308
F	317.31580	-6.83052
G	275.17819	-6.24286
H	509.57596	-10.72030
I	1114.17310	-37.18764
J	18.08666	-6.86518
K	94.16003	-5.52628
L	602.97382	-13.48760
M	470.51947	-9.76433
N	1129.48877	-38.11967
O	1169.50598	-40.62129
P	430.16336	-8.89147
Q	22.93244	-6.77219
R	983.84583	-29.76406
S	1086.57837	-35.53379
T	1353.57532	-53.42228
U	116.51624	-8.57173
V	167.72719	-5.44338
W	225.42348	-5.78933
X	56.45148	-6.01590
Y	225.59877	-5.73878
Z	11.55436	-7.26386
@	2853.20068	-227.53818

Table A.20: Classification matrix for training sample : polyalphabetic position 2, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	22.45	22	0	0	0	6	0	0	0	3	0	0	0	0	6
B	0.00	0	0	0	0	0	21	33	0	0	0	0	0	1	0
C	46.53	0	0	47	12	0	1	0	10	0	0	0	7	6	0
D	21.28	0	0	32	20	0	1	1	16	0	0	0	9	5	0
E	86.24	0	0	0	0	94	0	0	0	0	0	0	0	0	0
F	39.08	0	0	1	1	0	34	15	2	0	0	0	0	6	0
G	29.90	0	0	2	0	0	28	29	2	0	0	0	1	3	0
H	22.45	0	0	23	12	0	6	0	22	0	0	0	6	14	0
I	4.55	11	0	1	0	1	0	0	0	4	0	0	0	0	5
J	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	1	6	0	0	0	0	0	0	0
L	12.12	0	0	44	12	0	0	0	15	0	0	0	12	6	0
M	18.09	0	0	15	10	0	3	2	18	0	0	0	3	17	0
N	8.57	15	0	4	0	1	0	0	0	4	0	0	0	0	9
O	21.90	20	0	2	0	2	0	0	0	9	0	0	0	0	7
P	0.00	0	0	9	6	0	13	2	22	0	0	0	4	25	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	50.51	4	0	9	0	0	0	0	0	4	0	0	0	0	4
S	17.82	10	0	3	1	2	0	0	0	3	0	0	0	0	5
T	49.45	17	0	0	0	9	0	0	0	2	0	0	0	0	1
U	27.03	0	0	7	9	0	11	2	22	0	0	0	2	25	0
V	10.58	0	0	0	0	0	11	30	0	0	0	0	0	0	0
W	0.00	0	0	0	0	0	15	33	0	0	0	0	0	4	0
X	0.00	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Y	33.66	0	0	0	0	0	25	28	0	0	0	0	0	1	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	6	0	0	0
@	98.02	0	0	0	0	2	0	0	0	0	0	0	0	0	0
TOTAL	23.19	99	0	199	83	117	170	182	129	29	0	0	44	113	37

Table A.20 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	22.45	12	0	0	4	12	33	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	7	2	0	0	32	0	0	96
C	46.53	0	0	0	12	2	0	4	0	0	0	0	0	0	101
D	21.28	0	0	0	4	0	0	6	0	0	0	0	0	0	94
E	86.24	0	0	0	0	0	14	0	0	0	0	0	0	1	109
F	39.08	0	0	0	0	0	0	27	0	0	0	1	0	0	87
G	29.90	0	0	0	0	0	0	18	1	0	0	13	0	0	97
H	22.45	0	0	0	3	0	0	11	1	0	0	0	0	0	98
I	4.55	23	0	0	20	16	7	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	99	0	0	1	0	0	100
K	0.00	0	0	0	0	0	0	0	50	0	0	34	0	0	91
L	12.12	0	0	0	7	0	0	3	0	0	0	0	0	0	99
M	18.09	0	0	0	1	1	0	23	1	0	0	0	0	0	94
N	8.57	11	0	0	20	24	17	0	0	0	0	0	0	0	105
O	21.90	23	0	0	15	11	16	0	0	0	0	0	0	0	105
P	0.00	0	0	0	0	0	0	19	1	0	0	1	0	0	102
Q	0.00	0	0	0	0	0	0	0	98	0	0	4	0	0	102
R	50.51	9	0	0	50	17	2	0	0	0	0	0	0	0	99
S	17.82	15	0	0	32	18	12	0	0	0	0	0	0	0	101
T	49.45	13	0	0	2	2	45	0	0	0	0	0	0	0	91
U	27.03	0	0	0	0	0	0	30	1	0	0	2	0	0	111
V	10.58	0	0	0	0	0	0	0	11	0	0	52	0	0	104
W	0.00	0	0	0	0	0	0	9	0	0	0	32	0	0	93
X	0.00	0	0	0	0	0	0	0	84	0	0	20	0	0	105
Y	33.66	0	0	0	0	0	0	12	1	0	0	34	0	0	101
Z	0.00	0	0	0	0	0	0	0	102	0	0	0	0	0	102
@	98.02	0	0	0	0	0	0	0	0	0	0	0	0	99	101
TOTAL	23.19	106	0	0	170	103	146	169	452	0	0	226	0	100	2674

Table A.21: Classification Matrix for Validating Sample : Polyalphabetic Position 2, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	27.45	28	0	0	0	4	0	0	0	2	0	0	0	0	3
B	0.00	0	0	0	1	0	22	33	1	0	0	0	0	1	0
C	48.48	0	0	48	12	0	0	0	15	0	0	0	6	3	0
D	26.42	0	0	41	28	0	1	0	16	0	0	0	9	3	0
E	85.71	0	0	0	0	78	0	0	0	0	0	0	0	0	0
F	25.66	0	0	1	2	0	29	17	5	0	0	0	0	8	0
G	24.27	0	0	0	0	0	39	25	2	0	0	0	0	5	0
H	24.51	0	0	24	20	0	6	1	25	0	0	0	4	7	0
I	8.93	11	0	3	0	1	0	0	0	10	0	0	0	0	12
J	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	2	3	0	0	0	0	0	0	0
L	8.91	0	0	41	21	0	0	0	15	0	0	0	9	5	1
M	12.26	0	0	13	9	0	8	1	27	0	0	0	6	13	0
N	10.53	14	0	1	0	3	0	0	0	5	0	0	0	0	10
O	25.26	20	0	0	0	1	0	0	0	5	0	0	0	0	4
P	0.00	0	0	6	6	0	9	2	22	0	0	0	2	16	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	48.51	2	0	4	2	1	0	0	0	2	0	0	0	0	9
S	8.08	19	0	1	0	0	0	0	0	4	0	0	0	0	10
T	49.54	20	0	0	0	18	0	0	0	2	0	0	0	0	4
U	23.60	0	0	3	12	0	8	3	18	0	0	0	3	20	0
V	16.67	0	0	0	0	0	1	23	0	0	0	0	0	0	0
W	0.00	0	0	0	0	0	22	34	0	0	0	0	0	2	0
X	0.00	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Y	29.29	0	0	0	0	0	26	33	0	0	0	0	0	1	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	22.12	114	0	186	113	106	173	176	146	30	0	0	39	84	53

Table A.21 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	27.45	23	0	0	4	5	33	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	9	4	0	0	33	0	0	104
C	48.48	0	0	0	10	0	0	5	0	0	0	0	0	0	99
D	26.42	0	0	0	5	0	0	3	0	0	0	0	0	0	106
E	85.71	0	0	0	0	0	12	0	0	0	0	0	0	1	91
F	25.66	0	0	0	0	0	0	44	0	0	0	7	0	0	113
G	24.27	0	0	0	0	0	0	25	0	0	0	7	0	0	103
H	24.51	0	0	0	3	1	0	11	0	0	0	0	0	0	102
I	8.93	21	0	0	27	13	14	0	0	0	9	0	0	0	112
J	0.00	0	0	0	0	0	0	0	97	0	0	3	0	0	100
K	0.00	0	0	0	0	0	0	0	63	0	0	41	0	0	109
L	8.91	0	0	0	8	0	0	1	0	0	0	0	0	0	101
M	12.26	0	0	0	1	0	0	28	0	0	0	0	0	0	106
N	10.53	17	0	0	26	15	4	0	0	0	0	0	0	0	95
O	25.26	24	0	0	13	10	18	0	0	0	0	0	0	0	95
P	0.00	0	0	0	0	0	0	32	1	0	0	2	0	0	98
Q	0.00	0	0	0	0	0	0	0	93	0	0	5	0	0	98
R	48.51	9	0	0	49	21	2	0	0	0	0	0	0	0	101
S	8.08	19	0	0	32	8	6	0	0	0	0	0	0	0	99
T	49.54	6	0	0	3	2	54	0	0	0	0	0	0	0	109
U	23.60	0	0	0	0	0	0	21	0	0	0	1	0	0	89
V	16.67	0	0	0	0	0	0	0	16	0	0	56	0	0	96
W	0.00	0	0	0	0	0	0	15	0	0	0	34	0	0	107
X	0.00	0	0	0	0	0	0	0	83	0	0	11	0	0	95
Y	29.29	0	0	0	0	0	0	9	1	0	0	29	0	0	99
Z	0.00	0	0	0	0	0	0	0	97	0	0	1	0	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	22.12	119	0	0	181	75	143	203	455	0	0	230	0	100	2726

Table A.22: Coefficients of linear discriminant functions for character groups : polyalphabetic, position 2, 3 Variables

No.	Variable	RELFREQ	INFOCONT	CTOENTRO	CONSTANT
1	A	-1615.60046	2.15113	1520.26291	90.19623
2	B	-1372.85901	2.91210	853.69861	25.15195
3	C	-2167.42017	2.20063	1475.33972	55.09113
4	D	-2138.55737	2.24622	1427.05957	50.86934
5	E	-365.00711	2.13370	1133.13464	116.07045
6	F	-1731.65857	2.60026	1088.92383	32.54883
7	G	-1569.99158	2.73605	984.47791	28.90313
8	H	-2059.36597	2.31963	1356.04199	16.18557
9	I	-1891.50195	2.10486	1580.60596	82.73922
10	J	-91.89346	2.75563	84.08834	12.62255
11	K	-701.54041	3.12418	446.53665	18.13773
12	L	-2161.35327	2.21575	1457.98779	53.30216
13	M	-2003.77649	2.37050	1307.41821	13.28042
14	N	-1856.27661	2.11248	1570.34741	83.16597
15	O	-1796.95349	2.12034	1560.10454	85.18299
16	P	-1956.91064	2.41494	1262.58159	40.66850
17	Q	-128.98384	2.77963	106.08884	12.72988
18	R	-2059.64941	2.08914	1600.08105	76.27932
19	S	-1908.35925	2.10889	1575.07275	80.81212
20	T	-1454.77148	2.18246	1478.95105	91.12531
21	U	-1935.05359	2 43471	1244.35144	39.66751
22	V	-1149.72607	3.12033	714.35382	22.04452
23	W	-1406.86328	2.87613	875.36237	25.63545
24	X	-413.81720	3.59073	279.26657	17.20878
25	Y	-1406.66113	2.87847	875.37134	25.59610
26	Z	-49.27899	2.07461	51.90605	10.49232
27	[illegible]	3860.98804	3.52509	188.50330	239.42941

Table A.23: Classification matrix for training sample : polyalphabetic position 2, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	21.43	21	0	0	0	5	0	0	0	6	0	0	0	0	7
B	0.00	0	0	0	0	0	16	23	0	0	0	3	0	0	0
C	47.52	0	0	48	10	0	1	0	10	0	0	0	8	4	0
D	19.15	0	0	32	18	0	2	0	15	0	0	0	10	6	0
E	84.40	0	0	0	0	92	0	0	0	0	0	0	0	0	0
F	39.08	0	0	1	1	0	34	16	2	0	0	0	0	4	0
G	13.56	0	0	2	0	0	25	18	2	0	0	1	1	2	0
H	16.33	0	0	24	11	0	7	2	16	0	0	1	6	16	0
I	10.23	11	0	1	0	1	0	0	0	9	0	0	0	0	3
J	0.00	0	0	0	0	0	0	0	0	0	0	9	0	0	0
K	49.45	0	0	0	0	0	1	3	0	0	0	45	0	0	0
L	13.13	0	0	44	12	0	0	0	11	0	0	0	13	9	0
M	20.21	0	0	15	9	0	7	2	13	0	0	1	4	19	0
N	4.76	13	0	3	0	1	0	0	0	9	0	0	0	0	5
O	20.95	21	0	2	0	2	0	0	0	11	0	0	0	0	6
P	0.00	0	0	9	6	0	13	4	15	0	0	1	4	28	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	7	0	0	0
R	51.52	4	0	9	0	0	0	0	0	7	0	0	0	0	4
S	13.86	10	0	3	1	1	0	0	0	7	0	0	0	0	4
T	48.35	17	0	0	0	8	0	0	0	3	0	0	0	0	0
U	21.62	0	0	7	9	0	19	2	21	0	0	2	2	22	0
V	36.54	0	0	0	0	0	6	12	0	0	0	15	0	0	0
W	0.00	0	0	0	0	0	13	19	0	0	0	0	0	2	0
X	45.71	0	0	0	0	0	0	0	0	0	0	32	0	0	0
Y	19.80	0	0	0	0	0	23	22	0	0	0	4	0	1	0
Z	54.90	0	0	0	0	0	0	0	0	0	0	4	0	0	0
@	97.03	0	0	0	0	3	0	0	0	0	0	0	0	0	0
TOTAL	28.16	97	0	200	77	113	167	123	106	52	0	125	48	113	29

Table A 23 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	21.43	11	0	0	5	9	34	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	5	22	0	1	26	0	0	96
C	47.52	0	0	0	12	2	0	6	0	0	0	0	0	0	101
D	19.15	0	0	0	5	0	0	5	0	0	0	1	0	0	94
E	84.40	0	0	0	0	0	16	0	0	0	0	0	0	1	109
F	39.08	0	0	0	0	0	0	22	1	0	0	6	0	0	87
G	18.56	0	0	0	0	0	0	11	11	0	0	24	0	0	97
H	16.33	0	0	0	3		0	12	0	0	0	0	0	0	98
I	10.23	22	0	0	20	14	7	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	0	0	52	0	39	0	100
K	49.45	0	0	0	0	0	0	0	22	0	13	3	4	0	91
L	13.13	0	0	0	7	0	0	3	0	0	0	0	0	0	99
M	20.21	0	0	0	1	1	0	21	0	0	0	1	0	0	94
N	4.76	13	0	0	24	20	17	0	0	0	0	0	0	0	105
O	20.95	22	0	0	16	10	15	0	0	0	0	0	0	0	105
P	0.00	0	0	0	0	0	0	19	1	0	0	1	0	0	102
Q	0.00	0	0	0	0	0	0	0	2	0	53	2	38	0	102
R	51.52	8	0	0	51	14	2	0	0	0	0	0	0	0	99
S	13.86	15	0	0	34	14	12	0	0	0	0	0	0	0	101
T	48.35	15	0	0	2	2	44	0	0	0	0	0	0	0	91
U	21.62	0	0	0	0	0	0	24	0	0	0	3	0	0	111
V	36.54	0	0	0	0	0	0	0	38	0	2	31	0	0	104
W	0.00	0	0	0	0	0	0	7	24	0	0	28	0	0	93
X	45.71	0	0	0	0	0	0	0	14	0	48	1	10	0	105
Y	19.80	0	0	0	0	0	0	4	27	0	0	20	0	0	101
Z	54.90	0	0	0	0	0	0	0	0	0	42	0	56	0	102
@	97.03	0	0	0	0	0	0	0	0	0	0	0	0	98	101
TOTAL	28.16	106	0	0	180	86	147	139	162	0	211	147	147	99	2674

Table A.24: Classification matrix for validating sample : polyalphabetic position 2, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	27.45	28	0	0	0	3	0	0	0	2	0	0	0	0	4
B	0.00	0	0	0	1	0	16	21	1	0	0	6	0	1	0
C	45.45	0	0	45	11	0	0	0	15	0	0	0	7	2	0
D	26.42	0	0	40	28	0	0	1	12	0	0	0	11	5	0
E	85.71	0	0	0	0	78	0	0	0	0	0	0	0	0	0
F	24.78	0	0	1	2	0	28	18	4	0	0	0	0	5	0
G	27.18	0	0	0	0	0	29	28	2	0	0	1	0	4	0
H	20.59	0	0	24	17	0	7	2	21	0	0	0	7	10	0
I	15.18	13	0	3	0	1	0	0	0	17	0	0	0	0	8
J	0.00	0	0	0	0	0	0	0	0	0	0	12	0	0	0
K	52.29	0	0	0	0	0	1	2	0	0	0	57	0	0	0
L	7.92	0	0	41	21	0	0	0	14	0	0	0	8	4	1
M	13.21	0	0	15	9	0	13	2	25	0	0	0	4	14	0
N	8.42	13	0	1	0	3	0	0	0	10	0	0	0	0	8
O	27.37	17	0	0	0	0	0	0	0	5	0	0	0	0	5
P	0.00	0	0	6	6	0	17	6	18	0	0	1	2	16	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	5	0	0	0
R	49.50	3	0	3	2	0	0	0	0	4	0	0	0	0	7
S	8.08	17	0	0	0	0	0	0	0	7	0	0	0	0	8
T	51.38	21	0	0	0	14	0	0	0	4	0	0	0	0	2
U	24.72	0	0	3	12	0	7	5	13	0	0	0	4	21	0
V	45.83	0	0	0	0	0	1	7	0	0	0	19	0	0	0
W	0.00	0	0	0	0	0	18	19	0	0	0	2	0	1	0
X	43.16	0	0	0	0	0	0	0	0	0	0	33	0	0	0
Y	33.33	0	0	0	0	0	23	17	0	0	0	3	0	0	0
Z	44.90	0	0	0	0	0	0	0	0	0	0	4	0	0	0
@	98.99	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	28.69	112	0	182	109	100	160	128	125	49	0	143	43	83	43

Table A.24 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	27.45	22	0	0	4	5	34	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	5	23	0	1	29	0	0	104
C	45.45	0	0	0	13	0	0	6	0	0	0	0	0	0	99
D	26.42	0	0	0	6	0	0	3	0	0	0	0	0	0	106
E	85.71	0	0	0	0	0	13	0	0	0	0	0	0	0	91
F	24.78	0	0	0	0	0	0	39	6	0	0	10	0	0	113
G	27.18	0	0	0	0	0	0	18	5	0	0	16	0	0	103
H	20.59	0	0	0	4	0	0	10	0	0	0	0	0	0	102
I	15.18	21	0	0	29	9	11	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	2	0	59	0	27	0	100
K	52.29	0	0	0	0	0	0	0	29	0	16	3	1	0	109
L	7.92	0	0	0	9	0	0	3	0	0	0	0	0	0	101
M	13.21	0	0	0	1	0	0	23	0	0	0	0	0	0	106
N	8.42	16	0	0	27	13	4	0	0	0	0	0	0	0	95
O	27.37	26	0	0	14	9	19	0	0	0	0	0	0	0	95
P	0.00	0	0	0	0	0	0	23	1	0	0	2	0	0	98
Q	0.00	0	0	0	0	0	0	0	3	0	56	0	34	0	98
R	49.50	9	0	0	50	21	2	0	0	0	0	0	0	0	101
S	8.08	20	0	0	33	8	6	0	0	0	0	0	0	0	99
T	51.38	7	0	0	3	2	56	0	0	0	0	0	0	0	109
U	24.72	0	0	0	0	0	0	22	1	0	0	1	0	0	89
V	45.83	0	0	0	0	0	0	0	44	0	2	22	1	0	96
W	0.00	0	0	0	0	0	0	9	24	0	0	34	0	0	107
X	43.16	0	0	0	0	0	0	0	5	0	41	2	14	0	95
Y	33.33	0	0	0	0	0	0	4	19	0	0	33	0	0	99
Z	44.90	0	0	0	0	0	0	0	1	0	49	0	44	0	98
@	98.99	0	0	0	0	0	0	0	0	0	0	0	0	98	99
TOTAL	28.69	121	0	0	193	67	145	165	163	0	224	152	121	98	2726

Table A.25: Coeficients of linear discriminant functions for character groups : polyalphabetic, position 2, 20 variables

No.	Variable	A	B	C
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	9534.50879	245047.96875	241323.96875
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	0.78335	-8.19585	2.60735
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	1411.19592	1724.10168	2300.83618
9	RCTOJEN3	-495.02203	-95.40870	15.47173
10	CCTOJEN3	-36.73038	-69.20622	-60.02985
11	RWCENTR3	-1.76204	0.31898	-3.04563
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-11746.26953	-15253.42480	-2[illegible]417.32617
14	RWCPROSS	8.45005	15.38803	16.27246
15	RCTOJENT	24.93772	-577.94116	-584.62244
16	RWCENTRO	15.04282	28.17582	26.08281
17	WINCOLSS	-12435.27148	-14439.15625	-24076.48438
18	CWCPROSS	4.08697	8.49207	-6.67679
19	CCTOJENT	249.02499	-229.72400	-400.74954
20	CWCENTRO	23.06535	9.77569	17.35027
	CONSTANT	-126.29819	-46.12092	-78.53065

Table A.25 (continued)

No.	Variable	D	E	F
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	286449.90625	76566.57813	286889.25069
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	4.21717	[illegible]	-3.75478
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2510.0441[illegible]	[illegible]79.88989	2115.39[illegible]
9	RCTOJEN3	48.93684	9.87782	102.25452
10	CCTOJEN3	-101.54601	-170.33069	-40.74755
11	RWCENTR3	-0.28979	-6.08940	1.42773
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-32570.30469	27075.12500	-20877.48047
14	RWCPROSS	16.69216	9.64931	16.42749
15	RCTOJENT	-869.02765	182.24333	-612.35229
16	RWCENTRO	20.34258	15.65040	25.50337
17	WINCOLSS	-23675.59570	-68358.63281	-22182.19531
18	CWCPROSS	-9.32615	11.44166	0.62608
19	CCTOJENT	-289.35529	98.88066	-376.99484
20	CWCENTRO	18.35372	24.06197	12.19105
	CONSTANT	-75.28715	-177.25356	-55.24427

Table A.25 (continued)

No.	Variable	G	H	I
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	257263.00000	292382.96875	3447.08789
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	-4.51348	-1.26433	2.10367
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	1910.53088	2396.59595	1540.37976
9	RCTOJEN3	-136.86412	-17.63165	-900.37915
10	CCTOJEN3	-46.74520	102.71537	80.01486
11	RWCENTR3	3.21740	3.83387	-0.13800
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-21393.18945	-31478.21289	-4476.07227
14	RWCPROSS	17.59057	9.30001	6.82287
15	RCTOJENT	-625.17188	-793.42932	49.90362
16	RWCENTRO	22.76063	17.63566	14.55668
17	WINCOLSS	-13716.24512	-22878.42578	-11596.38965
18	CWCPROSS	0.92384	3.76530	4.37021
19	CCTOJENT	-233.24449	-491.80273	419.58701
20	CWCENTRO	13.95642	11.17244	25.08985
	CONSTANT	-51.61558	-64.59805	-132.36867

Table A.25 (continued)

No.	Variable	J	K	L
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	25973.23242	151983.70313	212568.17188
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	1.12044	7.53516	2.29163
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	149.86021	1027.93188	2170.32788
9	RCTOJEN3	28.50325	-176.27681	35.89701
10	CCTOJEN3	1.71234	-34.08051	-205.18922
11	RWCENTR3	-1.83233	5.23028	-3.88568
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-3520.92017	-8574.37500	-33351.66406
14	RWCPROSS	13.84863	16.48057	17.36229
15	RCTOJENT	-162.65352	-318.85403	-751.93182
16	RWCENTRO	9.30232	13.90686	27.74223
17	WINCOLSS	1323.13342	-2641.09131	-14274.57617
18	CWCPROSS	3.81704	-4.65476	-6.92609
19	CCTOJENT	29.55605	-91.14555	-62.55613
20	CWCENTRO	1.36842	10.96647	20.85175
	CONSTANT	-12.34112	-28.16427	-83.62402

Table A.25 (continued)

No.	Variable	M	N	O
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	269120.90625	221775.00000	74545.46875
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	2.81929	4.88986	3.40285
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2322.90112	2109.34253	1550.16309
9	RCTOJEN3	-79.33803	383.15347	-71.76266
10	CCTOJEN3	-90.40010	-8.56635	-199.60017
11	RWCENTR3	-2.93733	-8.12322	-6.04622
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-26665.79297	-9580.60254	-14050.58691
14	RWCPROSS	16.92326	25.71817	16.27043
15	RCTOJENT	-677.10022	6.18632	-75.15700
16	RWCENTRO	27.67566	24.95573	19.99345
17	WINCOLSS	-19686.87695	-53715.01172	-32557.12500
18	CWCPROSS	-7.05989	-19.10619	0.03386
19	CCTOJENT	-308.05560	-1163.81018	-12.73118
20	CWCENTRO	17.38465	16.37544	23.32295
	CONSTANT	-69.26191	-111.54741	-121.46066

Table A.25 (continued)

No.	Variable	P	Q	R
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	248224.32813	23942.70508	158318.17188
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	-0.09609	-6.87888	0.14607
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2147.58911	146.33076	2001.32825
9	RCTOJEN3	-56.84218	9.81111	209.26573
10	CCTOJEN3	-136.88481	-22.27656	-261.15564
11	RWCENTR3	-2.52536	-1.33199	-6.65139
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-23592.84766	-1998.27869	-18378.93750
14	RWCPROSS	19.21308	28.84329	19.69191
15	RCTOJENT	-638.94623	-57.27362	-287.14993
16	RWCENTRO	29.83784	8.12168	24.16068
17	WINCOLSS	-16271.58301	-205.64603	-33147.92188
18	CWCPROSS	-5.62871	-3.05277	-7.75875
19	CCTOJENT	-251.05684	-22.68412	-381.03152
20	CWCENTRO	17.82907	1.30205	19.71518
	CONSTANT	-69.33815	-13.39964	-101.72494

Table A.25 (continued)

No.	Variable	S	T	U
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	102352.94531	160480.32813	218194.39063
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	-4.38636	7.22983	-4.25505
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	1764.72131	1904.27637	2061.77979
9	RCTOJEN3	307.18790	108.64310	-81.79709
10	CCTOJEN3	-157 60258	218.48138	-168.81491
11	RWCENTR3	-8.75556	-8.63856	-8.11390
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	171.66830	-25735.74023	-28670.01758
14	RWCPROSS	20.70733	12.42546	28.37205
15	RCTOJENT	-848.75330	-666.79590	-802.49255
16	RWCENTRO	26.40531	23.45274	36.01734
17	WINCOLSS	-27756.66211	-26152.37891	-5145.02979
18	CWCPROSS	-2.90672	-6.95277	-10.18073
19	CCTOJENT	81.11006	-289.88391	18.70912
20	CWCENTRO	22.45966	20.09746	22.45018
	CONSTANT	-114.45261	-117.77831	-77.04524

Table A.25 (continued)

No.	Variable	V	W	X
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	212022.48438	250786.95313	82767.68750
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0.00000
6	CWCPROS3	1.89520	-5.14303	0.15579
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	1564.90698	1767.26392	553.56726
9	RCTOJEN3	-264.00107	-136.39192	-111.47751
10	CCTOJEN3	-84.63590	-46.49168	-41.01757
11	RWCENTR3	8.63135	3.45195	4.36975
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-18832.66211	-16202.64355	-5435.59375
14	RWCPROSS	25.04599	9.83050	4.05003
15	RCTOJENT	-398.36414	-556.48981	-270.98776
16	RWCENTRO	8.87285	24.29119	11.81544
17	WINCOLSS	-5943.85645	-15628.77441	829.60254
18	CWCPROSS	-11.00099	10.56762	18.09612
19	CCTOJENT	-121.77821	-247.50342	79.95732
20	CWCENTRO	15.59515	7.76970	2.71275
	CONSTANT	-41.28481	-44.96212	-20.41317

Table A.25 (continued)

No.	Variable	Y	Z	$\bar{\omega}$
1	RELFREQ	0.00000	0.00000	0.00000
2	ROWCOLSS	240498.26563	11431.50098	1034762.87500
3	WINROWS3	0.00000	0.00000	0.00000
4	WINCOLS3	0.00000	0.00000	0.00000
5	RWCPROS3	0.00000	0.00000	0 00000
6	CWCPROS3	-7.45835	4.07488	-12.67056
7	INFOCONT	0 00000	0.00000	0.00000
8	CTOENTRO	1846.95779	69.04932	2729.78857
9	RCTOJEN3	-214.25258	0.94349	248.30547
10	CCTOJEN3	-78.08704	10.21724	-639 35333
11	RWCENTR3	4.83386	-0.27589	-4 05328
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	-29157.35352	-1772.95239	-17685.86719
14	RWCPROSS	34.72898	8.01517	25.17219
15	RCTOJENT	-646.08411	-78.28469	1415 79028
16	RWCENTRO	12.85919	4.71125	1.68898
17	WINCOLSS	-4122.37402	785.86853	-191390.85938
18	CWCPROSS	-13.70343	3.06745	-6.60851
19	CCTOJENT	-63.11975	20.37779	-2020.27820
20	CWCENTRO	19.68484	0.78822	24.82429
	CONSTANT	-52.71482	-10.50505	-423.80396

Table A.26: Classification matrix for training sample : polyalphabetic position 2, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	53.06	52	0	0	0	0	0	0	0	20	0	0	0	0	0
B	26.04	0	25	0	0	0	13	13	0	0	1	3	0	1	0
C	52.48	0	0	53	4	0	1	1	1	0	0	0	14	14	1
D	80.85	0	0	5	76	0	0	3	2	0	0	0	3	3	0
E	94.50	2	0	0	0	103	0	0	0	1	0	0	0	0	0
F	58.62	0	3	0	4	0	51	12	6	0	0	0	0	1	0
G	46.39	0	7	2	3	0	14	45	6	0	0	3	0	5	0
H	81.63	0	0	2	1	0	6	4	80	0	0	0	0	3	0
I	82.95	9	0	0	0	1	0	0	0	73	0	0	1	0	0
J	32.00	0	0	0	0	0	0	0	0	0	32	6	0	0	0
K	41.76	0	5	0	0	0	1	4	0	0	4	38	0	0	0
L	64.65	0	0	8	2	0	0	0	0	0	0	0	64	6	0
M	26.60	0	0	20	3	0	6	1	3	0	0	1	7	25	0
N	86.67	0	0	4	0	0	0	0	0	0	0	0	0	0	91
O	72.38	10	0	0	0	1	0	0	0	0	0	0	2	0	0
P	46.08	0	0	12	2	0	4	4	1	0	0	1	7	16	0
Q	27.45	0	2	0	0	0	0	0	0	0	29	0	0	0	0
R	75.76	0	0	1	1	0	0	0	0	0	0	0	5	0	7
S	86.14	0	0	0	5	2	0	0	0	0	0	0	2	0	0
T	74.73	8	0	0	0	0	0	0	0	0	0	0	0	0	3
U	72.07	0	1	1	1	0	1	1	0	0	0	2	13	4	0
V	62.50	0	0	0	0	0	1	8	0	0	2	6	0	0	0
W	45.16	0	19	0	0	0	12	14	2	0	0	2	0	2	0
X	45.71	0	2	0	0	0	0	0	0	0	16	20	0	0	0
Y	68.32	0	1	0	1	0	0	8	0	0	0	1	0	0	0
Z	54.90	0	0	0	0	0	0	0	0	0	30	2	0	0	0
Φ	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	61.71	81	65	108	103	107	110	118	101	94	114	85	118	80	102

Table A.26 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	53.06	20	0	0	2	0	4	0	0	0	0	0	0	0	98
B	26.04	0	3	0	0	0	0	2	0	32	2	1	0	0	96
C	52.48	0	6	0	6	0	0	0	0	0	0	0	0	0	101
D	80.85	0	1	0	0	0	0	0	1	0	0	0	0	0	94
E	94.50	2	0	0	0	0	1	0	0	0	0	0	0	0	109
F	58.62	0	6	0	0	0	0	0	0	3	0	1	0	0	87
G	46.39	0	3	0	0	0	0	0	2	5	0	2	0	0	97
H	81.63	0	0	0	0	0	0	0	0	1	1	0	0	0	98
I	82.95	3	0	0	0	0	1	0	0	0	0	0	0	0	88
J	32.00	0	0	6	0	0	0	0	0	0	17	0	39	0	100
K	41.76	0	0	1	0	0	0	0	18	5	9	2	4	0	91
L	64.65	0	6	0	4	1	0	8	0	0	0	0	0	0	99
M	26.60	0	19	0	0	0	0	9	0	0	0	0	0	0	94
N	86.67	0	0	0	9	0	1	0	0	0	0	0	0	0	105
O	72.38	76	0	0	9	1	6	0	0	0	0	0	0	0	105
P	46.08	0	47	0	0	0	6	8	0	0	0	0	0	0	102
Q	27.45	0	0	28	0	0	0	0	0	0	5	0	38	0	102
R	75.76	4	0	0	75	1	5	0	0	0	0	0	0	0	99
S	86.14	0	0	0	4	87	1	0	0	0	0	0	0	0	101
T	74.73	4	0	0	7	1	68	0	0	0	0	0	0	0	91
U	72.07	0	6	0	0	0	0	80	0	0	0	1	0	0	111
V	62.50	0	0	1	0	0	0	0	65	0	0	21	0	0	104
W	45.16	0	0	0	0	0	0	0	0	42	0	0	0	0	93
X	45.71	0	0	1	0	0	0	0	2	6	48	0	10	0	105
Y	68.32	0	0	0	0	0	0	0	20	1	0	69	0	0	101
Z	54.90	0	0	6	0	0	0	0	0	0	8	0	56	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	61.71	109	97	43	116	91	87	107	108	95	90	97	147	101	2674

Table A.27: Classification matrix for validating sample : polyalphabetic position 2, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	55.88	57	0	0	0	1	0	0	0	14	0	0	0	0	0
B	28.85	0	30	1	0	0	11	8	1	0	1	1	0	5	0
C	57.58	0	0	57	4	0	2	0	3	0	0	0	6	9	1
D	69.81	0	0	12	74	0	0	0	1	0	0	0	6	9	0
E	95.60	0	0	0	0	87	0	0	0	1	0	0	0	0	0
F	37.17	0	4	2	2	0	42	27	9	0	0	1	0	14	0
G	48.54	0	3	0	1	0	19	50	5	0	0	0	0	6	0
H	84.31	0	0	1	4	0	4	5	86	0	0	0	0	1	0
I	74.11	19	0	0	0	0	0	0	0	83	0	0	2	0	0
J	31.00	0	0	0	0	0	0	0	0	0	31	6	0	0	0
K	40.37	0	4	0	0	0	0	10	0	0	7	44	0	0	0
L	71.29	0	0	4	5	0	0	0	1	0	0	0	72	4	0
M	26.42	0	0	18	7	0	3	0	1	0	0	0	9	28	0
N	95.79	0	0	2	0	0	0	0	0	0	0	0	0	0	91
O	67.37	13	0	0	1	0	0	0	0	3	0	0	0	0	0
P	43.88	0	0	10	2	0	6	6	0	0	1	1	8	12	0
Q	23.47	0	0	0	0	0	0	0	0	0	29	3	0	0	0
R	72.28	1	0	1	1	0	0	0	0	1	0	0	3	2	10
S	87.88	0	0	0	0	0	0	0	0	0	0	0	1	0	0
T	77.98	5	0	0	1	0	0	0	0	2	0	0	0	0	1
U	75.28	0	0	0	1	0	0	2	0	0	0	0	10	7	0
V	60.42	0	1	0	0	0	0	6	0	0	1	13	0	0	0
W	43.93	0	18	0	0	0	16	17	1	0	0	4	0	4	0
X	42.11	0	3	0	0	0	0	0	0	0	14	9	0	0	0
Y	80.81	0	0	0	1	0	1	4	0	0	0	3	0	0	0
Z	44.90	0	0	0	0	0	0	0	0	0	30	1	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	60.23	95	63	108	104	88	104	135	108	104	114	86	117	101	103

Table A.27 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	55.88	22	0	0	1	4	3	6	0	0	0	0	0	0	102
B	28.85	0	2	0	0	0	0	1	3	35	3	2	0	0	104
C	57.58	0	10	0	7	5	0	0	0	0	0	0	0	0	99
D	69.81	0	0	0	3	0	0	0	0	0	0	1	0	0	106
E	95.60	3	0	0	0	0	0	0	0	0	0	0	0	0	91
F	37.1.	0	5	0	0	0	0	0	4	3	0	0	0	0	113
G	48.54	0	7	0	0	0	6	3	3	6	0	0	0	0	103
H	84.31	0	0	0	0	0	0	0	0	1	0	0	0	0	102
I	74.11	5	0	0	0	2	1	0	0	0	0	0	0	0	112
J	31.00	0	0	12	0	0	6	0	0	0	24	0	27	0	100
K	40.37	0	0	3	0	0	0	0	17	6	16	1	1	0	109
L	71.29	1	2	0	5	0	0	7	0	0	0	0	0	0	101
M	26.42	0	26	0	0	0	0	14	0	0	0	0	0	0	106
N	95.79	0	0	0	1	0	1	0	0	0	0	0	0	0	95
O	67.37	64	0	0	8	0	6	0	0	0	0	0	0	0	95
P	43.88	0	43	0	0	0	0	8	0	0	0	1	0	0	98
Q	23.47	0	0	23	0	0	0	0	2	0	7	0	34	0	98
R	72.28	7	0	0	73	0	2	0	0	0	0	0	0	0	101
S	87.88	1	0	0	7	87	3	0	0	0	0	0	0	0	99
T	77.98	11	0	0	3	[illegible]	85	0	0	0	0	0	0	0	109
U	75.28	0	2	0	0	0	0	67	0	0	0	0	6	0	89
V	60.42	0	0	1	0	0	0	0	58	1	0	14	1	0	96
W	43.93	0	0	0	0	0	0	0	0	47	0	0	0	0	107
X	42.11	0	0	6	0	0	0	0	4	5	40	0	14	0	95
Y	80.81	0	0	0	0	0	0	1	9	0	0	80	0	0	99
Z	44.90	0	0	7	0	0	0	0	0	1	15	0	44	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	60.23	114	97	52	108	94	101	101	100	105	105	99	121	99	2726

Table A.28: Coefficients of linear discrimination for character groups, polyalphabetic, position 3, 1 variable

Group	Coefficients of RELFREQ	CONSTANT
A	1308.77686	-48.78096
B	223.34047	-5.81713
C	665.98199	-15.30924
D	590.54480	-12.83637
E	1813.39099	-90.91305
F	340.13272	-7.15545
G	271.83231	-6.13403
H	539.12366	-11.35086
I	1161.21704	-39.09519
J	21.09636	-6.86810
K	106.28788	-5.58480
L	624.67578	-13.91945
M	493.67239	-10.19163
N	1174.61621	-39.91764
O	1200.53882	-41.52723
P	449.08655	-9.19419
Q	21.24591	-6.76977
R	1038.12085	-31.95929
S	1118.62292	-36.50912
T	1367.28162	-52.98643
U	434.97339	-8.85870
V	168.70221	-5.43010
W	226.83833	-5.76661
X	57.72976	-6.01733
Y	239.33272	-5.87077
Z	13.32243	-7.26494
@	2933.44165	-233.64592

Table A.29: Classification matrix for training sample : polyalphabetic position 3, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	19.39	19	0	0	0	9	0	0	0	4	0	0	0	0	3
B	0.00	0	0	0	0	0	15	57	0	0	0	0	0	1	0
C	45.54	0	0	46	18	0	0	0	8	0	0	0	9	1	0
D	17.02	0	0	31	16	0	0	0	24	0	0	0	10	6	0
E	84.40	2	0	0	0	92	0	0	0	0	0	0	0	0	0
F	26.44	0	0	1	1	0	23	18	2	0	0	0	0	5	0
G	44.33	0	0	0	0	0	28	43	0	0	0	0	1	2	0
H	17.35	0	0	22	15	0	1	4	17	0	0	0	10	11	0
I	5.68	15	0	1	0	0	0	0	0	5	0	0	0	0	1
J	0.00	0	0	0	0	0	1	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	1	14	0	0	0	0	0	0	0
L	12.12	0	0	47	14	0	1	0	12	0	0	0	12	4	0
M	15.96	0	0	12	9	0	8	4	19	0	0	0	7	15	0
N	0.95	18	0	2	0	1	0	0	0	3	0	0	0	0	1
O	24.76	19	0	2	0	3	0	0	0	1	0	0	0	0	3
P	0.00	0	0	8	12	0	12	4	11	0	0	0	2	24	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R	44.44	7	0	10	0	1	0	0	0	2	0	0	1	0	1
S	19.80	8	0	4	1	2	0	0	0	2	0	0	0	0	2
T	39.56	22	0	0	0	14	0	0	0	0	0	0	0	0	4
U	32.43	0	0	5	8	0	12	5	23	0	0	0	4	17	0
V	15.38	0	0	0	0	0	5	51	0	0	0	0	0	0	0
W	16.13	0	0	0	0	0	10	52	1	0	0	0	0	2	0
X	0.00	0	0	0	0	0	0	6	0	0	0	0	0	0	0
Y	0.00	0	0	0	0	0	21	62	1	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	21.80	110	0	191	94	122	138	320	118	17	0	0	56	88	15

Table A.29 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	19.39	16	0	0	6	8	33	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	6	5	12	0	0	0	0	96
C	45.54	1	0	0	14	0	0	4	0	0	0	0	0	0	101
D	17.02	0	0	0	3	0	0	4	0	0	0	0	0	0	94
E	84.40	1	0	0	0	0	12	0	0	0	0	0	0	2	109
F	26.44	0	0	0	0	0	0	36	0	1	0	0	0	0	87
G	44.33	0	0	0	0	0	0	18	0	5	0	0	0	0	97
H	17.35	0	0	0	2	0	0	16	0	0	0	0	0	0	98
I	5.68	17	0	0	21	16	12	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	98	1	0	0	0	0	100
K	0.00	0	0	0	0	0	0	1	53	22	0	0	0	0	91
L	12.12	0	0	0	4	0	0	5	0	0	0	0	0	0	99
M	15.96	0	0	0	1	1	0	18	0	0	0	0	0	0	94
N	0.95	15	0	0	30	12	23	0	0	0	0	0	0	0	105
O	24.76	26	0	0	14	18	19	0	0	0	0	0	0	0	105
P	0.00	0	0	0	1	0	0	27	0	1	0	0	0	0	102
Q	0.00	0	0	0	0	0	0	0	100	2	0	0	0	0	102
R	44.44	9	0	0	44	22	2	0	0	0	0	0	0	0	99
S	19.80	20	0	0	30	20	12	0	0	0	0	0	0	0	101
T	39.56	6	0	0	6	3	36	0	0	0	0	0	0	0	91
U	32.43	0	0	0	0	0	0	36	0	1	0	0	0	0	111
V	15.38	0	0	0	0	0	0	1	16	31	0	0	0	0	104
W	16.13	0	0	0	0	0	0	10	3	15	0	0	0	0	93
X	0.00	0	0	0	0	0	0	0	92	7	0	0	0	0	105
Y	0.00	0	0	0	0	0	0	8	3	6	0	0	0	0	101
Z	0.00	0	0	0	0	0	0	0	102	0	0	0	0	0	102
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	101	101
TOTAL	21.80	111	0	0	176	100	149	190	472	104	0	0	0	103	2674

Table A.30: Classification matrix for validating sample : polyalphabetic position 3, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.51	25	0	0	0	6	0	0	0	4	0	0	0	0	1
B	0.00	0	0	0	0	0	12	68	0	0	0	0	0	1	0
C	38.38	0	0	38	14	0	1	0	14	0	0	0	6	11	0
D	24.53	0	0	37	26	0	0	0	14	0	0	0	9	11	0
E	84.62	1	0	0	0	77	0	0	0	0	0	0	0	0	0
F	19.47	0	0	1	3	0	22	27	8	0	0	0	1	13	0
G	46.60	0	0	0	[illegible]	0	28	48	0	0	0	0	0	5	0
H	18.63	0	0	28	17	0	3	2	19	0	0	0	9	12	0
I	2.68	22	0	6	0	0	0	0	0	3	0	0	0	0	4
J	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K	0.00	0	0	0	0	0	1	15	0	0	0	0	0	0	0
L	8.91	0	0	41	18	0	1	0	11	0	0	0	9	7	0
M	19.81	0	0	12	10	0	8	3	21	0	0	0	7	21	0
N	2.11	15	0	3	0	1	0	0	0	4	0	0	0	0	2
O	23.16	17	0	2	0	3	0	0	0	6	0	0	0	0	5
P	0.00	0	0	3	5	0	15	5	19	0	0	0	2	19	0
Q	0.00	0	0	0	0	0	0	1	0	0	0	0	0	0	0
R	44.55	4	0	12	0	0	0	0	0	2	0	0	0	0	4
S	19.19	10	0	3	0	0	0	0	0	2	0	0	0	0	8
T	55.96	17	0	0	0	15	0	0	0	0	0	0	0	0	0
U	29.21	0	0	3	6	0	14	4	22	0	0	0	4	10	0
V	19.79	0	0	0	0	0	3	43	0	0	0	0	0	0	0
W	7.48	0	0	0	0	0	17	69	0	0	0	0	0	2	0
X	0.00	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Y	0.00	0	0	0	0	0	19	56	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	21.61	111	0	189	99	102	144	344	128	21	0	0	47	112	24

Table A.30 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	24.51	14	0	0	2	11	39	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	10	4	9	0	0	0	0	104
C	38.38	0	0	0	9	0	0	6	0	0	0	0	0	0	99
D	24.53	0	0	0	3	0	0	6	0	0	0	0	0	0	106
E	84.62	0	0	0	0	0	12	0	0	0	0	0	0	1	91
F	19.47	0	0	0	0	0	0	37	0	1	0	0	0	0	113
G	46.60	0	0	0	0	0	0	19	0	3	0	0	0	0	103
H	18.63	0	0	0	0	0	0	12	0	0	0	0	0	0	102
I	2.68	19	0	0	26	24	8	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	100	0	0	0	0	0	100
K	0.00	0	0	0	0	0	0	0	68	25	0	0	0	0	109
L	8.91	1	0	0	4	0	0	9	0	0	0	0	0	0	101
M	19.81	0	0	0	0	0	0	24	0	0	0	0	0	0	106
N	2.11	14	0	0	32	15	9	0	0	0	0	0	0	0	95
O	23.16	22	0	0	9	15	16	0	0	0	0	0	0	0	95
P	0.00	0	0	0	0	0	0	29	0	1	0	0	0	0	98
Q	0.00	0	0	0	0	0	0	0	94	3	0	0	0	0	98
R	44.55	16	0	0	45	15	3	0	0	0	0	0	0	0	101
S	19.19	20	0	0	27	19	10	0	0	0	0	0	0	0	99
T	55.96	8	0	0	4	4	61	0	0	0	0	0	0	0	109
U	29.21	0	0	0	0	0	0	26	0	0	0	0	0	0	89
V	19.79	0	0	0	0	0	0	1	19	30	0	0	0	0	96
W	7.48	0	0	0	0	0	0	4	7	8	0	0	0	0	107
X	0.00	0	0	0	0	0	0	0	89	3	0	0	0	0	95
Y	0.00	0	0	0	0	0	0	13	2	9	0	0	0	0	99
Z	0.00	0	0	0	0	0	0	0	98	0	0	0	0	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	21.61	114	0	0	161	103	158	196	481	92	0	0	0	100	2726

Table A.31: Coefficients of linear discriminant functions for character groups : polyalphabetic, position 3, 3 Variables

No.	Variable	RELFREQ	INFOCONT	CTOENTRO	CONSTANT
1	A	-1785.48218	2.22654	1632.17078	-94.70445
2	B	-1483.91309	2.95834	917.73871	-26.27851
3	C	-2353.12988	2.25870	1593.38879	-59.41434
4	D	-2324.38501	2.30251	1539.61487	-54.49320
5	E	-508.71698	2.49521	1233.06519	-120.06177
6	F	-1897.27478	2.63461	1190.37805	-35.27285
7	G	-1668.28894	2.80960	1037.42932	-29.63959
8	H	-2240.34375	2.36781	1469.77820	-49.99491
9	I	-2050.08057	2.18292	1692.63623	-87.97156
10	J	-126.52832	2.85810	105.21648	-13.06583
11	K	-841.09491	3.51765	527.91125	-19.65038
12	L	-2343.64307	2.28015	1567.17273	-56.82028
13	M	-2173.93213	2.42015	1412.09253	-46.45369
14	N	-2015.24414	2.19074	1681.55957	-88.24403
15	O	-1981.93042	2.19390	1677.74512	-89.66698
16	P	-2113.18213	2.47074	1357.78564	-43.31286
17	Q	-129.54623	3.08270	109.09658	-13.95833
18	R	-2206.86646	2.16898	1710.02759	-81.70007
19	S	-2081.84253	2.18626	1687.03333	-85.10440
20	T	-1659.40955	2.25104	1597.25659	-97.25290
21	U	-2099.69556	2.48434	1343.55969	-42.43236
22	V	-1225.59607	3.17922	757.09283	-22.61890
23	W	-1474.97449	2.97557	915.07886	-26.23123
24	X	-450.63565	3.52455	299.54465	-16.84653
25	Y	-1552.07837	2.90390	960.98761	-27.37689
26	Z	-63.64581	2.57818	65.66998	-12.19959
27	@	3912.81689	3.54159	-474.40625	-245.27840

Table A.32: Classification matrix for training sample : polyalphabetic position 3, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	20.41	20	0	0	0	7	0	0	0	7	0	0	0	0	1
B	0.00	0	0	0	0	0	12	35	0	0	0	4	0	0	0
C	45.54	0	0	46	18	0	2	0	8	0	0	0	9	1	0
D	17.02	0	0	30	16	0	1	0	22	0	0	0	10	6	0
E	84.40	2	0	0	0	92	0	0	0	0	0	0	0	0	0
F	33.33	0	0	1	1	0	29	18	2	0	0	0	0	3	0
G	25.77	0	0	0	0	0	28	25	0	0	0	1	1	2	0
H	15.31	0	0	22	15	0	3	3	15	0	0	0	9	13	0
I	6.82	15	0	1	0	0	0	0	0	6	0	0	0	0	2
J	0.00	0	0	0	0	0	0	1	0	0	0	4	0	0	0
K	41.76	0	0	0	0	0	0	2	0	0	0	38	0	0	0
L	12.12	0	0	45	13	0	2	0	13	0	0	0	12	4	0
M	12.77	0	0	11	8	0	14	3	19	0	0	0	8	12	0
N	0.95	16	0	1	0	1	0	0	0	4	0	0	0	0	1
O	24.76	21	0	2	0	2	0	0	0	5	0	0	0	0	0
P	0.00	0	0	7	12	0	17	6	10	0	0	0	2	22	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	5	0	0	0
R	45.45	4	0	10	0	1	0	0	0	4	0	0	1	0	1
S	17.82	11	0	4	1	2	0	0	0	4	0	0	0	0	0
T	41.76	20	0	0	0	12	0	0	0	4	0	0	0	0	0
U	23.42	0	0	5	7	0	23	4	20	0	0	0	4	19	0
V	43.27	0	0	0	0	0	5	15	0	0	0	16	0	0	0
W	24.73	0	0	0	0	0	13	16	1	0	0	3	0	1	0
X	46.67	0	0	0	0	0	0	2	0	0	0	31	0	0	0
Y	5.94	0	0	0	0	0	17	36	1	0	0	2	0	0	0
Z	45.10	0	0	0	0	0	0	0	0	0	0	4	0	0	0
@	98.02	0	0	0	0	2	0	0	0	0	0	0	0	0	0
TOTAL	27.41	109	0	185	91	119	166	166	111	34	0	108	56	83	5

Table A.32 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	20.41	17	0	0	8	6	32	0	0	0	0	0	0	0	98
B	0.00	0	0	0	0	0	0	3	17	19	1	5	0	0	96
C	45.54	1	0	0	14	0	0	2	0	0	0	0	0	0	101
D	17.02	0	0	0	4	0	0	5	0	0	0	0	0	0	94
E	84.4	1	0	0	0	0	13	0	0	0	0	0	0	1	109
F	33.33	0	0	0	0	0	0	28	2	3	0	0	0	0	87
G	25.77	0	0	0	0	0	0	10	12	15	0	3	0	0	97
H	15.31	0	0	0	3	0	0	14	0	1	0	0	0	0	98
I	6.82	17	0	0	22	15	10	0	0	0	0	0	0	0	88
J	0.00	0	0	0	0	0	0	0	1	0	57	0	37	0	100
K	41.76	0	0	0	0	0	0	1	30	2	15	2	1	0	91
L	12.12	0	0	0	6	0	0	4	0	0	0	0	0	0	99
M	12.77	0	0	0	2	1	0	15	0	1	0	0	0	0	94
N	0.95	17	0	0	31	12	22	0	0	0	0	0	0	0	105
O	24.76	26	0	0	16	15	18	0	0	0	0	0	0	0	105
P	0.00	0	0	0	2	0	0	22	1	0	0	1	0	0	102
Q	0.00	0	0	0	0	0	0	0	2	0	62	0	33	0	102
R	45.45	11	0	0	45	20	2	0	0	0	0	0	0	0	99
S	17.82	20	0	0	32	18	9	0	0	0	0	0	0	0	101
T	41.76	8	0	0	6	3	38	0	0	0	0	0	0	0	91
U	23.42	0	0	0	0	0	0	26	2	1	0	0	0	0	111
V	43.27	0	0	0	0	0	0	0	45	17	1	5	0	0	104
W	24.73	0	0	0	0	0	0	7	20	23	2	7	0	0	93
X	46.67	0	0	0	0	0	0	0	9	1	49	0	13	0	105
Y	5.94	0	0	0	0	0	0	4	16	18	1	6	0	0	101
Z	45.10	0	0	0	0	0	0	0	0	0	52	0	46	0	102
@	98.02	0	0	0	0	0	0	0	0	0	0	0	0	99	101
TOTAL	27.41	118	0	0	191	90	144	141	157	101	240	29	130	100	2674

Table A.33: Classification matrix for validating sample : polyalphabetic position 3, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.51	25	0	0	0	6	0	0	0	6	0	0	0	0	0
B	0.00	0	0	0	0	0	18	30	0	0	0	5	0	1	0
C	37.37	0	0	37	13	0	2	0	14	0	0	0	7	9	0
D	23.58	0	0	38	25	0	1	0	15	0	0	0	7	11	0
E	84.62	1	0	0	0	77	0	0	0	0	0	0	0	0	0
F	27.43	0	0	1	3	0	31	23	7	0	0	0	1	11	0
G	29.13	0	0	0	0	0	27	30	0	0	0	0	0	4	0
H	16.67	0	0	27	17	0	6	3	17	0	0	0	9	13	0
I	8.04	19	0	6	0	0	0	0	0	9	0	0	0	0	1
J	0.00	0	0	0	0	0	0	0	0	0	0	7	0	0	0
K	39.45	0	0	0	0	0	0	2	0	0	0	43	0	0	0
L	7.92	0	0	42	17	0	1	0	11	0	0	0	8	8	0
M	17.92	0	0	12	10	0	13	4	19	0	0	0	7	19	0
N	1.05	15	0	3	0	1	0	0	0	7	0	0	0	0	1
O	24.21	16	0	2	0	2	0	0	0	12	0	0	0	0	1
P	0.00	0	0	3	4	0	17	7	15	0	0	0	2	22	0
Q	0.00	0	0	0	0	0	0	1	0	0	0	11	0	0	0
R	44.55	4	0	12	0	0	0	0	0	6	0	0	0	0	0
S	17.17	9	0	2	0	0	0	0	0	11	0	0	0	0	2
T	59.63	16	0	0	0	11	0	0	0	1	0	0	0	0	0
U	25.84	0	0	3	6	0	19	4	18	0	0	0	4	11	0
V	48.96	0	0	0	0	0	1	11	0	0	0	19	0	0	0
W	17.76	0	0	0	0	0	15	39	0	0	0	5	0	2	0
X	48.42	0	0	0	0	0	0	0	0	0	0	34	0	0	0
Y	10.10	0	0	0	0	0	24	30	0	0	0	2	0	0	0
Z	48.98	0	0	0	0	0	0	0	0	0	0	2	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	28.03	105	0	188	95	97	175	184	116	52	0	128	45	111	5

Table A.33 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	24.51	16	0	0	3	9	37	0	0	0	0	0	0	0	102
B	0.00	0	0	0	0	0	0	2	19	21	0	8	0	0	104
C	37.37	0	0	0	10	0	0	7	0	0	0	0	0	0	99
D	23.58	0	0	0	4	0	0	5	0	0	0	0	0	0	106
E	84.62	0	0	0	0	0	13	0	0	0	0	0	0	0	91
F	27.43	0	0	0	0	0	0	23	2	10	0	1	0	0	113
G	29.13	0	0	0	0	0	0	16	7	14	0	5	0	0	103
H	16.67	0	0	0	1	0	0	9	0	0	0	0	0	0	102
I	8.04	21	0	0	31	17	8	0	0	0	0	0	0	0	112
J	0.00	0	0	0	0	0	0	0	0	0	62	0	31	0	100
K	39.45	0	0	0	0	0	0	0	27	8	22	3	4	0	109
L	7.92	1	0	0	4	0	0	9	0	0	0	0	0	0	101
M	17.92	0	0	0	0	0	0	22	0	0	0	0	0	0	106
N	1.05	13	0	0	34	12	9	0	0	0	0	0	0	0	95
O	24.21	23	0	0	10	12	17	0	0	0	0	0	0	0	95
P	0.00	0	0	0	0	0	0	25	2	1	0	0	0	0	98
Q	0.00	0	0	0	0	0	0	0	3	0	54	0	29	0	98
R	44.55	16	0	0	45	15	3	0	0	0	0	0	0	0	101
S	17.17	20	0	0	28	17	10	0	0	0	0	0	0	0	99
T	59.63	9	0	0	4	3	65	0	0	0	0	0	0	0	109
U	25.84	0	0	0	0	0	0	23	0	1	0	0	0	0	89
V	48.96	0	0	0	0	0	0	1	47	13	2	2	0	0	96
W	17.76	0	0	0	0	0	0	1	18	19	2	6	0	0	107
X	48.42	0	0	0	0	0	0	0	5	1	46	0	9	0	95
Y	10.10	0	0	0	0	0	0	4	14	15	0	10	0	0	99
Z	48.98	0	0	0	0	0	0	0	0	0	48	0	48	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	28.03	119	0	0	174	85	162	147	144	103	236	35	121	99	2726

Table A.34: Coeficients of Linear Discriminant Functions for Character Groups : Polyalphabetic, Position 3, 20 variables

No.	Variable	A	B	C
1	RELFREQ	-5092.24854	-6625.89111	-7539.19971
2	ROWCOLSS	80186.39844	447968.37500	441202.90625
3	WINROWS3	46485.66797	2845.16162	-326.02466
4	WINCOLS3	-5158.41309	-2886.97852	689.62067
5	RWCPROS3	10.43306	7.50972	11.24577
6	CWCPROS3	-5.24859	-5.98147	2.21007
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	1721.56458	2629.52686	3252.16138
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-4.89573	-14.99985	-9.85050
15	RCTOJENT	205.19395	-74.55418	59.67832
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	9.32018	20.07393	1.14651
19	CCTOJENT	346.15152	-43.79667	-137.00487
20	CWCENTRO	26.25520	15.18779	21.45585
	CONSTANT	-130.66689	-42.40907	-84.54173

Table A.34 (continued)

No.	Variable	D	E	F
1	RELFREQ	-7512.96582	-2462.00830	-7618.48242
2	ROWCOLSS	491177.25000	36203.34766	514881.75000
3	WINROWS3	-4496.65137	-17270.96289	1864.95056
4	WINCOLS3	1203.24402	9300.65234	-5216.64111
5	RWCPROS3	4.71382	10.66395	6.03990
6	CWCPROS3	4.35453	-8.74163	-1.92972
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	3444.29224	862.92743	3170.74585
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-1.50108	-12.29687	-9.13457
15	RCTOJENT	-224.15742	-315.21710	-72.37935
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	-2.91166	22.72359	9.42526
19	CCTOJENT	-15.31324	910.03467	-104.88348
20	CWCENTRO	21.86513	28.61839	16.54394
	CONSTANT	-81.93446	-173.37320	-55.90642

Table A.34 (continued)

No.	Variable	G	H	I
1	RELFREQ	-7080.57764	-7031.25977	-6561.72998
2	ROWCOLSS	480207 59375	512914.68750	134222.71875
3	WINROWS3	2643.51245	-2759.18677	76674.71094
4	WINCOLS3	-4054.24780	-11968.73242	-13073.34473
5	RWCPROS3	5.27401	4.11798	14.66341
6	CWCPROS3	-2.28894	-6.85814	-6.95253
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2870.05957	3493.96997	2047.37732
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0 00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-4.72272	-5.24904	-7.94550
15	RCTOJENT	-82.90034	-169.41626	84.65507
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	7.08083	11.78820	9.37269
19	CCTOJENT	-63.16313	-224.50345	526.00873
20	CWCENTRO	18.41632	16.11842	27.74748
	CONSTANT	-49.98328	-71.25595	-139.28526

Table A.34 (continued)

No.	Variable	J	K	L
1	RELFREQ	-1022.93005	-4330.80029	-8045.86572
2	ROWCOLSS	65604.29688	304998.34375	455975.46875
3	WINROWS3	-1033.66663	4453.19629	-10.36793
4	WINCOLS3	-184.44667	-4835.17236	3829.51733
5	RWCPROS3	9.91883	-7.79756	13.07884
6	CWCPROS3	0.95865	11.70165	3.52201
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	357.21494	1703.34131	3225.57080
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-1.25906	4.84483	-11.99774
15	RCTOJENT	-13 86803	-53.98373	-36.66420
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	7.56495	3.44247	0.48620
19	CCTOJENT	24.56989	-39.07122	58.83969
20	CWCENTRO	2.83869	12.19930	25.29860
	CONSTANT	-12.15248	-24.33109	-90.24912

Table A.34 (continued)

No.	Variable	M	N	O
1	RELFREQ	-7899.76807	-3752.08521	-5494.94971
2	ROWCOLSS	491618.90625	196771.18750	144249.71875
3	WINROWS3	3876.20361	-20996.94141	3603.67334
4	WINCOLS3	331.10898	3370.97900	9538.70703
5	RWCPROS3	12.72731	12.96183	12.46224
6	CWCPROS3	5.01228	-0.38016	-5.40980
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	3309.04028	2377.32495	1904.10352
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-13.93909	1.05451	-4.73330
15	RCTOJENT	-24.48735	395.82431	182.04628
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	1.12737	-6.89504	6.83479
19	CCTOJENT	-101.64040	-529.88861	333.49850
20	CWCENTRO	22.19914	21.62439	26.41669
	CONSTANT	-74.60025	-110.24218	-124.03643

Table A.34 (continued)

No.	Variable	P	Q	R
1	RELFREQ	-8045.29834	-858.24780	-5515.48193
2	ROWCOLSS	496560.43750	39068.35938	237710.51563
3	WINROWS3	3109.89063	1036.57190	-11420.48145
4	WINCOLS3	575.03113	1697.16260	14124.98730
5	RWCPROS3	9.32883	1.04244	11.98600
6	CWCPROS3	6.23704	-5.01257	-0.73650
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	3236.88916	229.09241	2491.32324
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-13.27708	19.45157	-4.54973
15	RCTOJENT	-11.52494	67.65196	176.92929
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	2.66483	3.69346	0.41199
19	CCTOJENT	-62.30359	-14.18200	-59.41484
20	CWCENTRO	22.33913	4.56633	24.12143
	CONSTANT	-72.16533	-14.29469	-105.90031

Table A.34 (continued)

No.	Variable	S	T	U
1	RELFREQ	-4606.05371	-3016.84302	-8322.33301
2	ROWCOLSS	270625.81250	158705.65625	497690.53125
3	WINROWS3	-18403.99414	6604.57080	6658.14795
4	WINCOLS3	4676.29932	-17149.12695	473.80667
5	RWCPROS3	11.80552	13.69554	18.05980
6	CWCPROS3	-2.84221	0.03596	1.62404
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2605.12183	2117.37817	3188.95752
9	RCTOJEN3	0 00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	-8.15760	-9.38387	-14.17884
15	RCTOJENT	-542.44373	-102.61370	-55.22809
16	RWCENTRO	0 00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	6.44145	3.62704	-0.25243
19	CCTOJENT	410.32925	-3.21367	57.98935
20	CWCENTRO	27.45439	25.59826	26.12461
	CONSTANT	-120.38746	-118.39574	-79 16724

Table A.34 (continued)

No.	Variable	V	W	X
1	RELFREQ	-5481.52441	-6509.16064	-2597.66455
2	ROWCOLSS	392263.25000	442809.75000	171611 75000
3	WINROWS3	5141.50449	2963.60376	2416.86890
4	WINCOLS3	-4444.95898	-3821.86011	-1652.84583
5	RWCPROS3	-7.70808	2.91368	-3.83218
6	CWCPROS3	6.45252	-3.62007	4.71712
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2282.16479	2611.62378	939.37061
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0 00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	20.30695	-13.66209	-5.69819
15	RCTOJENT	-90 86066	-75 25359	-59.94365
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	-9.34650	21.46569	21.35145
19	CCTOJENT	-88.22146	-53.37277	54 47816
20	CWCENTRO	18.30135	13.44182	6.31053
	CONSTANT	-40.19640	-40.63849	-16.71425

Table A.34 (continued)

No.	Variable	Y	Z	@
1	RELFREQ	-6408.01611	-700.80603	-5996.21680
2	ROWCOLSS	450965.03125	40512.11328	584860.06250
3	WINROWS3	2963.92700	168.84755	-53162.21875
4	WINCOLS3	-2326.79443	-1255.06885	38820.72266
5	RWCPROS3	4.55189	2.29269	13.64696
6	CWCPROS3	-11.72595	7.12432	-18.45299
7	INFOCONT	0.00000	0.00000	0.00000
8	CTOENTRO	2696.52148	212.09259	1076.98816
9	RCTOJEN3	0.00000	0.00000	0.00000
10	CCTOJEN3	0.00000	0.00000	0.00000
11	RWCENTR3	0.00000	0.00000	0.00000
12	CWCENTR3	0.00000	0.00000	0.00000
13	WINROWSS	0.00000	0.00000	0.00000
14	RWCPROSS	26.22813	1.48954	14.43631
15	RCTOJENT	-160.40169	4.88280	857.95612
16	RWCENTRO	0.00000	0.00000	0.00000
17	WINCOLSS	0.00000	0.00000	0.00000
18	CWCPROSS	-11.93842	5.74223	2.36825
19	CCTOJENT	-42.37393	23.92611	85.82717
20	CWCENTRO	22.24163	1.64712	25.88114
	CONSTANT	-55.96051	-11.60525	-329.14731

Table A.35: Classification matrix for training sample : polyalphabetic position 3, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	58.16	57	0	0	0	0	0	0	0	15	0	0	0	0	0
B	14.58	0	14	0	0	0	15	23	0	0	0	4	0	0	0
C	58.42	0	0	59	2	0	1	0	4	0	0	0	14	10	2
D	80.85	0	0	8	76	0	0	0	1	0	0	0	4	2	0
E	93.58	1	0	0	0	102	0	0	0	1	0	0	0	0	0
F	49.43	0	2	1	1	0	43	15	5	0	0	0	0	5	0
G	37.11	0	4	1	0	0	24	36	2	0	0	2	0	2	0
H	80.61	0	0	0	1	0	13	2	79	0	0	0	0	1	0
I	78.41	11	0	0	0	0	0	0	0	69	0	0	0	0	0
J	48.00	0	0	0	0	0	0	0	0	0	48	1	0	0	0
K	42.86	0	4	0	0	0	1	4	0	0	3	39	0	0	0
L	68.69	0	0	9	3	0	0	0	0	0	0	0	68	3	0
M	28.72	0	0	24	1	0	2	3	0	0	0	0	6	27	1
N	91.43	0	0	2	0	0	0	0	0	0	0	0	0	0	96
O	75.24	6	0	0	1	1	0	0	0	1	0	0	3	0	0
P	34.31	0	0	15	4	0	5	5	2	0	0	0	3	16	0
Q	41.18	0	0	0	0	0	0	0	0	0	29	2	0	0	0
R	64.65	0	0	6	0	0	0	0	0	0	0	0	5	0	10
S	85.15	0	0	0	3	0	0	0	0	0	0	0	2	0	0
T	75.82	4	0	0	1	0	0	0	0	1	0	0	0	0	1
U	61.26	0	0	0	2	0	0	4	1	0	0	0	22	6	0
V	53.85	0	2	0	1	0	1	10	0	0	0	9	0	1	0
W	43.01	0	18	0	0	0	15	8	3	0	0	3	0	0	0
X	35.24	0	2	0	0	0	0	0	0	0	23	21	0	0	0
Y	79.21	0	1	0	1	0	1	7	1	0	0	0	0	0	0
Z	41.18	0	0	0	0	0	0	0	0	0	44	0	0	0	0
@	99.01	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	60.25	79	47	125	97	104	121	117	98	87	147	81	127	73	110

Table A.35 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	58.16	19	0	0	2	1	4	0	0	0	0	0	0	0	98
B	14.58	0	1	0	0	0	0	0	0	39	0	0	0	0	96
C	58.42	0	4	0	4	0	1	0	0	0	0	0	0	0	101
D	80.85	0	3	0	0	0	0	0	0	0	0	0	0	0	94
E	93.58	3	0	0	0	2	0	0	0	0	0	0	0	0	109
F	49.43	0	5	0	0	0	0	0	1	9	0	0	0	0	87
G	37.11	0	13	0	0	0	0	2	4	7	0	0	0	0	97
H	80.61	0	0	0	0	0	0	0	0	2	0	0	0	0	98
I	78.41	6	0	0	0	1	1	0	0	0	0	0	0	0	88
J	48.00	0	0	6	0	0	0	0	0	1	14	0	30	0	100
K	42.86	0	0	6	0	0	0	0	17	4	11	1	1	0	91
L	68.69	0	2	0	3	1	0	10	0	0	0	0	0	0	99
M	28.72	0	20	0	0	0	0	10	0	0	0	0	0	0	94
N	91.43	0	0	0	5	0	2	0	0	0	0	0	0	0	105
O	75.24	79	0	0	8	2	4	0	0	0	0	0	0	0	105
P	34.31	0	35	0	2	0	0	14	0	0	0	1	0	0	102
Q	41.18	0	6	42	0	0	0	0	0	0	4	0	25	0	102
R	64.65	5	0	0	64	4	5	0	0	0	0	0	0	0	99
S	85.15	2	0	0	1	86	7	0	0	0	0	0	0	0	101
T	75.82	4	0	0	8	3	69	0	0	0	0	0	0	0	91
U	61.26	0	7	0	0	0	0	68	1	0	0	0	0	0	111
V	53.85	0	1	0	0	0	0	0	56	1	2	20	0	0	104
W	43.01	0	0	0	0	0	0	0	5	40	1	0	0	0	93
X	35.24	0	0	6	0	0	0	0	2	5	37	0	9	0	105
Y	79.21	0	0	0	0	0	0	1	8	0	1	80	0	0	101
Z	41.18	0	0	6	0	0	0	0	0	0	10	0	42	0	102
@	99.01	0	0	0	0	0	0	0	0	0	0	0	0	100	101
TOTAL	60.25	118	91	66	97	100	93	105	94	108	80	102	107	100	2674

Table A.36: Classification Matrix for Validating Sample : Polyalphabetic Position 3, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	53.92	55	0	0	0	0	0	0	0	17	0	0	0	0	0
B	16.35	0	17	0	0	0	17	22	1	0	0	4	0	0	0
C	47.47	0	0	47	5	0	1	1	1	0	0	0	11	13	1
D	77.36	0	0	10	82	0	0	0	4	0	0	0	4	3	0
E	94.51	2	0	0	0	86	0	0	0	0	0	0	0	0	0
F	38.94	0	4	2	0	0	44	22	11	0	0	1	0	14	0
G	44.66	0	5	0	1	0	17	46	5	0	0	1	0	3	0
H	78.43	0	0	3	5	0	8	3	80	0	0	0	0	1	0
I	70.54	19	0	1	0	0	0	0	0	79	0	0	3	0	0
J	39.00	0	0	0	0	0	0	0	0	0	39	2	0	0	0
K	45.87	0	3	0	0	0	0	7	0	0	5	50	0	0	0
L	59.41	0	0	6	10	0	0	0	1	0	0	0	60	4	0
M	23.58	0	0	16	8	0	4	2	4	0	0	0	7	25	0
N	81.05	0	0	5	0	0	0	0	0	0	0	0	0	0	77
O	73.68	6	0	0	1	1	0	0	0	0	0	0	1	0	0
P	31.63	0	0	4	1	0	8	12	1	0	0	0	5	25	0
Q	39.80	0	0	0	0	0	0	0	0	0	24	2	0	0	0
R	72.28	1	0	4	3	0	0	0	0	0	0	0	5	0	4
S	91.92	0	0	0	1	0	0	0	0	0	0	0	1	0	0
T	80.73	4	0	2	0	0	0	0	0	1	0	0	0	0	1
U	64.04	0	0	0	1	0	1	2	0	0	0	0	19	7	0
V	54.17	0	2	0	0	0	1	4	0	0	1	11	0	6	0
W	39.25	0	24	0	1	0	17	12	1	0	0	4	0	1	0
X	44.21	0	2	0	0	0	0	0	0	0	14	11	0	0	0
Y	76.77	0	0	0	1	0	1	6	0	0	0	1	0	0	0
Z	44.90	0	0	0	0	0	0	0	0	0	40	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	58.36	87	57	100	119	87	119	139	109	97	123	87	116	96	83

Table A.36 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	53.92	21	0	0	1	0	8	0	0	0	0	0	0	0	102
B	16.35	0	1	0	0	0	0	0	0	41	1	0	0	0	104
C	47.47	0	11	0	6	0	0	2	0	0	0	0	0	0	99
D	77.36	0	2	0	0	0	0	0	0	0	0	1	0	0	106
E	94.51	3	0	0	0	0	0	0	0	0	0	0	0	0	91
F	38.94	0	7	0	0	0	0	0	1	7	0	0	0	0	113
G	44.66	0	13	0	0	0	0	0	2	8	0	2	0	0	103
H	78.43	0	0	0	0	0	0	0	0	2	0	0	0	0	102
I	70.54	8	0	0	0	0	2	0	0	0	0	0	0	0	112
J	39.00	0	0	18	0	0	0	0	0	0	12	0	29	0	100
K	45.87	0	0	5	0	0	0	0	24	3	8	1	3	0	109
L	59.41	0	3	0	2	1	0	14	0	0	0	0	0	0	101
M	23.58	0	27	0	0	0	0	13	0	0	0	0	0	0	106
N	81.05	0	0	0	12	0	1	0	0	0	0	0	0	0	95
O	73.68	70	0	0	7	6	3	0	0	0	0	0	0	0	95
P	31.63	0	31	0	0	0	0	11	0	0	0	0	0	0	98
Q	39.80	0	0	39	0	0	0	0	2	1	7	0	23	0	98
R	72.28	4	0	0	73	2	5	0	0	0	0	0	0	0	101
S	91.92	1	0	0	4	91	1	0	0	0	0	0	0	0	99
T	80.73	5	0	0	4	4	88	0	0	0	0	0	0	0	109
U	64.04	0	2	0	0	0	0	57	0	0	0	0	0	0	89
V	54.17	0	0	0	0	0	0	0	52	0	0	25	0	0	96
W	39.25	0	0	0	0	0	0	0	2	42	4	0	0	0	107
X	44.21	0	0	11	0	0	0	0	0	5	42	0	10	0	95
Y	76.77	0	0	0	0	0	0	3	10	1	0	76	0	0	99
Z	44.90	0	0	6	0	0	0	0	0	0	8	0	44	0	98
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	99	99
TOTAL	58.36	112	97	79	109	104	108	100	93	110	82	105	109	99	2726

APPENDIX B

Results from Quadratic Discriminant Analysis

Table B.1: Classification matrix for training sample : quadratic discriminant, monoalphabetic, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	27.00	27	0	0	0	3	0	0	0	0	0	0	0	0	7
B	0.00	0	0	0	0	0	5	27	0	0	0	1	0	0	0
C	31.00	0	0	31	10	0	0	0	9	0	0	0	33	0	0
D	19.00	0	0	11	19	0	0	0	19	0	0	0	38	1	0
E	94.00	0	0	0	0	94	0	0	0	0	0	0	0	0	0
F	39.00	0	0	0	1	0	39	24	1	0	0	0	0	1	0
G	33.00	0	0	0	0	0	28	33	0	0	0	0	0	0	0
H	18.00	0	0	4	13	0	2	1	18	0	0	0	31	2	0
I	0.00	12	0	0	0	0	0	0	0	0	0	0	0	0	17
J	0.00	0	0	0	0	0	0	0	0	0	0	4	0	0	0
K	60.00	0	0	0	0	0	0	1	0	0	0	60	0	0	0
L	36.00	0	0	18	22	0	0	0	15	0	0	0	36	1	0
M	2.00	0	0	2	12	0	8	1	22	0	0	0	9	2	0
N	26.00	18	0	0	0	0	0	0	0	0	0	0	0	0	26
O	32.00	18	0	1	0	0	0	0	0	0	0	0	0	0	18
P	0.00	0	0	3	4	0	15	5	14	0	0	0	4	2	0
Q	70.00	0	0	0	0	0	0	0	0	0	0	6	0	0	0
R	65.00	1	0	3	0	0	0	0	0	0	0	0	0	0	19
S	0.00	15	0	1	0	0	0	0	0	0	0	0	1	0	17
T	57.00	20	0	0	0	9	0	0	0	0	0	0	0	0	3
U	61.00	0	0	0	2	0	13	4	14	0	0	0	4	1	0
V	56.00	0	0	0	0	0	1	4	0	0	0	12	0	0	0
W	14.00	0	0	0	0	0	5	25	1	0	0	0	0	0	0
X	42.00	0	0	0	0	0	0	0	0	0	0	28	0	0	0
Y	25.00	0	0	0	0	0	9	33	0	0	0	1	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	33.59	111	0	74	83	106	125	158	113	0	0	112	156	10	107

Table B.1 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	27.00	26	0	0	1	0	36	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	3	28	18	0	18	0	0	100
C	31.00	0	0	0	5	0	0	12	0	0	0	0	0	0	100
D	19.00	0	0	0	3	0	0	9	0	0	0	0	0	0	100
E	94.00	0	0	0	0	0	6	0	0	0	0	0	0	0	100
F	39.00	0	0	0	0	0	0	25	1	1	0	7	0	0	100
G	33.00	0	0	0	0	0	0	5	9	9	0	16	0	0	100
H	18.00	0	0	0	1	0	[illegible]	28	0	0	0	0	0	0	100
I	0.00	26	0	0	37	0	8	0	0	0	0	0	0	0	100
J	0.00	0	0	71	0	0	0	0	0	0	25	0	0	0	100
K	60.00	0	0	2	0	0	0	1	25	1	10	0	0	0	100
L	36.00	0	0	0	4	0	0	4	0	0	0	0	0	0	100
M	2.00	0	0	0	2	0	0	42	0	0	0	0	0	0	100
N	26.00	20	0	0	30	0	6	0	0	0	0	0	0	0	100
O	32.00	32	0	0	15	0	16	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	52	0	0	0	1	0	0	100
Q	70.00	0	0	70	0	0	0	0	0	0	24	0	0	0	100
R	65.00	12	0	0	65	0	0	0	0	0	0	0	0	0	100
S	0.00	21	0	0	41	0	4	0	0	0	0	0	0	0	100
T	57.00	10	0	0	1	0	57	0	0	0	0	0	0	0	100
U	61.00	0	0	0	0	0	0	61	0	0	0	1	0	0	100
V	56.00	0	0	0	0	0	0	0	56	18	0	9	0	0	100
W	14.00	0	0	0	0	0	0	0	30	14	0	25	0	0	100
X	42.00	0	0	20	0	0	0	0	9	1	42	0	0	0	100
Y	25.00	0	0	0	0	0	0	2	15	15	0	25	0	0	100
Z	0.00	0	0	92	0	0	0	0	1	0	7	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	33.59	147	0	255	205	0	133	244	174	77	108	102	0	100	2700

Table B.2: Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.00	24	0	0	0	0	0	0	0	0	0	0	0	0	9
B	0.00	0	0	0	0	0	4	31	0	0	0	2	0	1	0
C	25.00	0	0	25	4	0	0	0	11	0	0	0	43	2	0
D	12.00	0	0	14	12	0	0	0	16	0	0	0	42	2	0
E	96.00	0	0	0	0	96	0	0	0	0	0	0	0	0	0
F	45.00	0	0	0	1	0	45	26	1	0	0	0	0	0	0
G	36.00	0	0	0	0	0	26	36	1	0	0	0	0	1	0
H	16.00	0	0	9	10	0	4	1	16	0	0	0	27	0	0
I	0.00	10	0	0	0	0	0	0	0	0	0	0	0	0	24
J	0.00	0	0	0	0	0	0	0	0	0	0	4	0	0	0
K	51.00	0	0	0	0	0	0	0	0	0	0	51	0	0	0
L	44.00	0	0	14	14	0	0	0	14	0	0	0	44	0	0
M	0.00	0	0	4	6	0	12	0	14	0	0	0	16	0	0
N	21.00	14	0	0	0	1	0	0	0	0	0	0	0	0	21
O	34.00	24	0	0	0	0	0	0	0	0	0	0	0	0	18
P	0.00	0	0	0	5	0	22	4	9	0	0	0	3	1	0
Q	67.00	0	0	0	0	0	0	0	0	0	0	5	0	0	0
R	71.00	0	0	3	0	0	0	0	0	0	0	0	0	0	15
S	0.00	5	0	2	0	0	0	0	0	0	0	0	0	0	21
T	64.00	15	0	0	0	10	0	0	0	0	0	0	0	0	4
U	53.00	0	0	0	2	0	14	7	19	0	0	0	3	1	0
V	61.00	0	0	0	0	0	1	3	0	0	0	10	0	0	0
W	16.00	0	0	0	0	0	12	19	0	0	0	1	0	0	0
X	39.00	0	0	0	0	0	0	0	0	0	0	35	0	0	0
Y	24.00	0	0	0	0	0	7	27	0	0	0	1	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	0	2	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	33.30	92	0	71	54	107	147	154	101	0	0	111	178	8	112

Table B.2 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	24.00	26	0	0	3	0	38	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	1	24	21	0	16	0	0	100
C	25.00	0	0	0	9	0	0	6	0	0	0	0	0	0	100
D	12.00	0	0	0	0	0	0	14	0	0	0	0	0	0	100
E	96.00	0	0	0	0	0	4	0	0	0	0	0	0	0	100
F	45.00	0	0	0	0	0	0	22	1	1	0	3	0	0	100
G	36.00	0	0	0	0	0	0	9	2	6	0	19	0	0	100
H	16.00	0	0	0	1	0	0	31	1	0	0	0	0	0	100
I	0.00	25	0	0	32	0	9	0	0	0	0	0	0	0	100
J	0.00	0	0	78	0	0	0	0	0	0	18	0	0	0	100
K	51.00	0	0	3	0	0	0	0	31	2	12	1	0	0	100
L	44.00	0	0	0	5	0	0	9	0	0	0	0	0	0	100
M	0.00	0	0	0	0	0	0	47	0	0	0	1	0	0	100
N	21.00	22	0	0	28	0	14	0	0	0	0	0	0	0	100
O	34.00	34	0	0	12	0	12	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	55	0	0	0	1	0	0	100
Q	67.00	0	0	67	0	0	0	0	4	0	24	0	0	0	100
R	71.00	10	0	0	71	0	1	0	0	0	0	0	0	0	100
S	0.00	31	0	0	35	0	6	0	0	0	0	0	0	0	100
T	64.00	5	0	0	2	0	64	0	0	0	0	0	0	0	100
U	53.00	0	0	0	0	0	0	53	0	0	0	1	0	0	100
V	61.00	0	0	0	0	0	0	0	61	14	1	10	0	0	100
W	16.00	0	0	0	0	0	0	3	28	16	0	21	0	0	100
X	39.00	0	0	19	0	0	0	0	6	0	39	1	0	0	100
Y	24.00	0	0	0	0	0	0	1	26	14	0	24	0	0	100
Z	0.00	0	0	89	0	0	0	0	0	0	9	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	33.30	153	0	256	198	0	148	251	184	74	103	98	0	100	2700

Table B.3: Classification matrix for training sample : quadratic discriminant, monoalphabetic, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	34.00	34	0	0	0	4	0	0	0	0	0	0	0	0	8
B	0.00	0	0	0	0	0	10	33	0	0	0	0	0	0	0
C	31.00	0	0	31	6	0	0	0	7	0	0	0	39	6	0
D	19.00	0	0	10	19	0	0	0	12	0	0	0	47	2	0
E	94.00	0	0	0	0	94	0	0	0	0	0	0	0	0	0
F	40.00	0	0	0	1	0	40	19	3	0	0	0	0	2	0
G	34.00	0	0	0	0	0	32	34	0	0	0	0	0	0	0
H	12.00	0	0	4	19	0	2	0	12	0	0	0	35	10	0
I	1.00	16	0	0	0	0	0	0	0	1	0	0	0	0	22
J	17.00	0	0	0	0	0	0	0	0	0	17	11	0	0	0
K	60.00	0	0	0	0	0	0	1	0	0	0	60	0	0	0
L	42.00	0	0	18	26	0	0	0	6	0	0	0	42	3	0
M	14.00	0	0	2	16	0	7	0	18	0	0	0	13	14	0
N	28.00	19	0	0	0	0	0	0	0	0	0	0	0	0	28
O	28.00	23	0	1	0	0	0	0	0	2	0	0	0	0	20
P	0.00	0	0	2	6	0	11	3	18	0	0	0	6	12	0
Q	0.00	0	0	0	0	0	0	0	0	0	12	11	0	0	0
R	58.00	2	0	3	0	0	0	0	0	2	0	0	0	0	22
S	0.00	17	0	1	0	0	0	0	0	0	0	0	1	0	21
T	57.00	21	0	0	0	11	0	0	0	0	0	0	0	0	3
U	45.00	0	0	0	5	0	10	3	15	0	0	0	4	18	0
V	48.00	0	0	0	0	0	1	9	0	0	0	8	0	0	0
W	21.00	0	0	0	0	0	9	36	1	0	0	0	0	0	0
X	41.00	0	0	0	0	0	0	0	0	0	1	46	0	0	0
Y	6.00	0	0	0	0	0	20	43	0	0	0	1	0	0	0
Z	21.00	0	0	0	0	0	0	0	0	0	35	3	0	0	0
Φ	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	31.52	132	0	72	98	109	142	181	92	5	65	140	187	67	124

Table B.3 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	34.00	17	0	0	0	0	37	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	3	17	27	0	10	0	0	100
C	31.00	0	0	0	5	0	0	6	0	0	0	0	0	0	100
D	19.00	0	0	0	3	0	0	7	0	0	0	0	0	0	100
E	94.00	0	0	0	0	0	6	0	0	0	0	0	0	0	100
F	40.00	0	0	0	0	0	0	33	0	1	0	1	0	0	100
G	34.00	0	0	0	0	0	0	13	3	13	0	5	0	0	100
H	12.00	0	0	0	1	0	0	17	0	0	0	0	0	0	100
I	1.00	22	0	0	30	0	9	0	0	0	0	0	0	0	100
J	17.00	0	0	0	0	0	0	0	0	0	67	0	5	0	100
K	60.00	0	0	0	0	0	0	1	29	3	6	0	0	0	100
L	42.00	0	0	0	4	0	0	1	0	0	0	0	0	0	100
M	14.00	0	0	0	2	0	0	28	0	0	0	0	0	0	100
N	28.00	20	0	0	25	0	8	0	0	0	0	0	0	0	100
O	28.00	28	0	0	10	0	16	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	42	0	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	2	0	67	0	8	0	100
R	58.00	13	0	0	58	0	0	0	0	0	0	0	0	0	100
S	0.00	20	0	0	36	0	4	0	0	0	0	0	0	0	100
T	57.00	7	0	0	1	0	57	0	0	0	0	0	0	0	100
U	45.00	0	0	0	0	0	0	45	0	0	0	0	0	0	100
V	48.00	0	0	0	0	0	0	0	48	28	0	6	0	0	100
W	21.00	0	0	0	0	0	0	0	21	21	0	12	0	0	100
X	41.00	0	0	0	0	0	0	0	10	1	41	0	1	0	100
Y	6.00	0	0	0	0	0	0	2	8	20	0	6	0	0	100
Z	21.00	0	0	0	0	0	0	0	1	0	40	0	21	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	31.52	127	0	0	175	0	137	198	139	114	221	40	35	100	2700

Table B.4: Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	32.00	32	0	0	0	0	0	0	0	0	0	0	0	0	9
B	0.00	0	0	0	0	0	12	33	1	0	0	1	0	0	0
C	23.00	0	0	23	7	0	0	0	9	0	0	0	47	1	0
D	20.00	0	0	14	20	0	0	0	10	0	0	0	45	7	0
E	96.00	0	0	0	0	96	0	0	0	0	0	0	0	0	0
F	51.00	0	0	0	2	0	51	15	0	0	0	0	0	2	0
G	40.00	0	0	0	1	0	34	40	1	0	0	0	0	1	0
H	8.00	0	0	9	11	0	3	1	8	0	0	0	34	13	0
I	1.00	16	0	0	0	0	0	0	0	1	0	0	0	0	29
J	20.00	0	0	0	0	0	0	0	0	0	20	7	0	0	0
K	54.00	0	0	0	0	0	0	1	0	0	0	54	0	0	0
L	55.00	0	0	13	7	0	0	0	11	0	0	0	55	4	0
M	13.00	0	0	4	9	0	8	1	12	0	0	0	18	13	0
N	22.00	18	0	0	0	1	0	0	0	1	0	0	0	0	[illegible]
O	30.00	29	0	0	0	0	0	0	0	0	0	0	0	0	1
P	0.00	0	0	0	4	0	14	4	9	0	0	0	6	14	0
Q	0.00	0	0	0	0	0	0	0	0	0	13	11	0	0	0
R	62.00	1	0	3	0	0	0	0	0	1	0	0	0	0	21
S	0.00	12	0	2	0	0	0	0	0	2	0	0	0	0	24
T	65.00	14	0	0	0	11	0	0	0	0	0	0	0	0	3
U	42.00	0	0	0	9	0	15	4	14	0	0	0	3	13	0
V	50.00	0	0	0	0	0	1	7	0	0	0	10	0	0	0
W	23.00	0	0	0	0	0	14	31	0	0	0	0	0	0	0
X	37.00	0	0	0	0	0	0	1	0	0	3	50	0	0	0
Y	8.00	0	0	0	0	0	12	37	0	0	0	1	0	0	0
Z	12.00	0	0	0	0	0	0	0	0	0	26	2	0	0	0
[illegible]	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	32.00	122	0	68	70	108	164	175	75	5	62	136	208	68	126

Table B.4 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	32.00	17	0	0	2	0	40	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	1	15	27	0	16	0	0	100
C	23.00	0	0	0	9	0	0	4	0	0	0	0	0	0	100
D	20.00	0	0	0	0	0	0	4	0	0	0	0	0	0	100
E	96.00	0	0	0	0	0	4	0	0	0	0	0	0	0	100
F	51.00	0	0	0	0	0	0	27	0	2	0	1	0	0	100
G	40.00	0	0	0	0	0	0	10	1	6	0	6	0	0	100
H	8.00	0	0	0	1	0	0	19	1	0	0	0	0	0	100
I	1.00	19	0	0	26	0	9	0	0	0	0	0	0	0	100
J	20.00	0	0	0	0	0	0	0	1	0	60	0	12	0	100
K	54.00	0	0	0	0	0	0	0	34	5	6	0	0	0	100
L	55.00	0	0	0	5	0	0	5	0	0	0	0	0	0	100
M	13.00	0	0	0	0	0	0	35	0	0	0	0	0	0	100
N	22.00	18	0	0	25	0	15	0	0	0	0	0	0	0	100
O	30.00	30	0	0	10	0	13	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	49	0	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	5	0	60	0	11	0	100
R	62.00	11	0	0	62	0	1	0	0	0	0	0	0	0	100
S	0.00	26	0	0	28	0	6	0	0	0	0	0	0	0	100
T	65.00	5	0	0	2	0	65	0	0	0	0	0	0	0	100
U	42.00	0	0	0	0	0	0	42	0	0	0	0	0	0	100
V	50.00	0	0	0	0	0	0	0	50	25	1	6	0	0	100
W	23.00	0	0	0	0	0	0	5	18	23	0	9	0	0	100
X	37.00	0	0	0	0	0	0	0	9	0	37	0	0	0	100
Y	8.00	0	0	0	0	0	0	3	16	23	0	8	0	0	100
Z	12.00	0	0	0	0	0	0	0	0	0	60	0	12	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	32.00	126	0	0	170	0	153	204	150	111	224	40	35	100	2700

Table B.5: Classification matrix for training sample : quadratic discriminant, monoalphabetic, 12 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	54.00	54	0	0	0	0	0	0	0	5	0	0	0	0	0
B	49.00	0	49	0	0	0	13	2	0	0	0	0	0	1	0
C	74.00	0	0	74	1	0	0	0	1	0	0	0	5	7	1
D	89.00	0	0	0	89	0	1	0	4	0	0	0	1	5	0
E	98.00	1	0	0	0	98	0	0	0	0	0	0	0	0	0
F	77.00	0	3	0	0	0	77	6	4	0	0	0	0	5	0
G	50.00	0	3	0	0	0	22	50	2	0	0	0	0	9	0
H	91.00	0	1	2	1	0	3	0	91	0	0	0	0	1	0
I	65.00	13	0	0	0	0	0	0	0	65	0	0	0	0	0
J	13.00	0	0	0	0	0	0	0	0	0	13	19	0	0	0
K	71.00	0	2	0	0	0	0	0	1	0	1	71	0	0	0
L	83.00	0	0	8	2	0	0	0	0	0	0	0	83	2	0
M	48.00	0	0	24	4	0	3	0	1	0	0	0	7	48	0
N	96.00	0	0	1	0	0	0	0	0	0	0	0	0	0	96
O	81.00	4	0	0	0	0	0	0	0	4	0	0	1	0	0
P	42.00	0	1	17	0	0	2	0	1	0	0	0	12	17	0
Q	74.00	0	0	0	0	0	0	0	0	0	1	6	0	0	0
R	84.00	0	0	3	0	0	0	0	0	0	0	0	2	0	2
S	95.00	0	0	0	0	0	0	0	0	1	0	0	1	0	0
T	98.00	0	0	0	0	0	0	0	0	0	0	0	1	0	0
U	84.00	0	0	1	0	0	0	1	0	0	0	0	10	1	0
V	77.00	0	0	0	0	0	1	2	0	0	0	3	0	0	0
W	61.00	0	14	0	0	0	17	5	0	0	0	1	0	0	0
X	63.00	0	5	0	0	0	0	0	0	0	0	25	0	0	0
Y	81.00	0	0	0	0	0	0	1	0	0	0	1	0	0	0
Z	38.00	0	0	0	0	0	0	0	0	0	10	3	0	0	0
α	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	71.70	72	78	130	97	98	139	67	105	75	25	129	123	96	99

Table B.5 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	54.00	40	0	0	1	0	0	0	0	0	0	0	0	0	100
B	49.00	0	4	0	0	0	0	0	1	30	0	0	6	0	100
C	74.00	0	9	0	2	0	0	0	0	0	0	0	0	0	100
D	89.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
E	98.00	0	0	0	0	1	0	0	0	0	0	0	0	0	100
F	77.00	0	1	0	0	0	0	0	0	4	0	0	0	0	100
G	50.00	0	5	0	0	0	0	0	1	8	0	0	0	0	100
H	91.00	0	1	0	0	0	0	0	0	0	0	0	0	0	100
I	65.00	18	0	0	1	3	0	0	0	0	0	0	0	0	100
J	13.00	0	0	20	0	0	0	0	0	0	37	0	11	0	100
K	71.00	0	0	1	0	0	0	0	14	5	5	0	0	0	100
L	83.00	0	2	0	0	0	0	3	0	0	0	0	0	0	100
M	48.00	0	10	0	0	0	0	3	0	0	0	0	0	0	100
N	96.00	0	0	0	3	0	0	0	0	0	0	0	0	0	100
O	81.00	81	0	0	4	0	6	0	0	0	0	0	0	0	100
P	42.00	0	42	0	0	0	0	8	0	0	0	0	0	0	100
Q	74.00	0	0	74	0	0	0	0	0	0	5	0	14	0	100
R	84.00	3	0	0	84	0	6	0	0	0	0	0	0	0	100
S	95.00	0	0	0	0	95	3	0	0	0	0	0	0	0	100
T	98.00	0	0	0	1	0	98	0	0	0	0	0	0	0	100
U	84.00	0	3	0	0	0	0	84	0	0	0	0	0	0	100
V	77.00	0	0	0	0	0	0	0	77	1	0	16	0	0	100
W	61.00	0	0	0	0	0	0	0	2	61	0	0	0	0	100
X	63.00	0	0	4	0	0	0	0	1	1	63	0	1	0	100
Y	81.00	0	0	0	0	0	0	0	17	0	0	81	0	0	100
Z	38.00	0	0	23	0	0	0	0	0	0	26	0	38	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	71.70	142	77	122	96	99	113	98	113	110	136	97	64	100	2700

Table B.6: Classification matrix for validating sample : quadratic discriminant, monoalphabetic, 12 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	50.00	50	0	0	0	0	0	0	0	8	0	0	0	0	0
B	48.00	0	48	0	0	0	9	13	0	0	0	1	0	0	0
C	75.00	0	0	75	2	0	0	0	1	0	0	0	9	1	2
D	82.00	0	0	4	82	0	0	0	4	0	0	0	3	7	0
E	99.00	1	0	0	0	99	0	0	0	0	0	0	0	0	0
F	72.00	0	4	0	0	0	72	5	3	0	0	0	0	6	0
G	47.00	0	5	0	2	0	22	47	3	0	0	0	0	7	0
H	86.00	0	0	0	4	0	8	1	86	0	0	1	0	0	0
I	58.00	10	0	0	0	0	0	0	0	58	0	0	0	0	0
J	20.00	0	0	0	0	0	0	0	0	0	20	13	0	0	0
K	68.00	0	3	0	0	0	0	3	0	0	1	68	0	0	0
L	82.00	0	0	2	0	0	0	0	0	0	0	0	82	2	0
M	51.00	0	0	20	3	0	7	0	2	0	0	0	7	51	0
N	94.00	0	0	0	0	0	0	0	0	0	0	0	0	0	94
O	86.00	0	0	0	0	0	0	0	0	9	0	0	0	0	0
P	51.00	0	2	7	0	0	6	3	0	0	0	0	3	16	0
Q	70.00	0	0	0	0	0	0	0	0	0	3	3	0	0	0
R	88.00	0	0	3	0	0	0	0	0	0	0	0	3	0	1
S	97.00	1	0	0	2	0	0	0	0	0	0	0	0	0	0
T	95.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
U	83.00	0	0	0	1	0	1	0	0	0	0	0	12	1	0
V	78.00	0	0	0	0	0	0	0	0	0	0	3	0	0	0
W	61.00	0	13	0	0	0	21	5	0	0	0	0	0	0	0
X	62.00	0	2	0	0	0	0	0	0	0	2	26	0	0	0
Y	86.00	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Z	22.00	0	0	0	0	0	0	0	0	0	13	3	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	70.78	62	77	111	96	99	146	77	99	75	39	119	119	91	97

Table B.6 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	50.00	35	0	0	3	1	3	0	0	0	0	0	0	0	100
B	48.00	0	2	0	0	0	0	0	1	26	0	0	0	0	100
C	75.00	0	7	0	3	0	0	0	0	0	0	0	0	0	100
D	82.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
E	99.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
F	72.00	0	6	0	0	0	0	0	0	4	0	0	0	0	100
G	47.00	0	11	0	0	0	0	0	0	2	0	1	0	0	100
H	86.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
I	58.00	30	0	0	1	1	0	0	0	0	0	0	0	0	100
J	20.00	0	0	19	0	0	0	0	0	0	31	0	17	0	100
K	68.00	0	0	2	0	0	0	0	11	6	5	1	0	0	100
L	82.00	0	0	0	4	0	0	10	0	0	0	0	0	0	100
M	51.00	0	10	0	0	0	0	0	0	0	0	0	0	0	100
N	94.00	0	0	0	4	0	2	0	0	0	0	0	0	0	100
O	86.00	86	0	0	1	1	3	0	0	0	0	0	0	0	100
P	51.00	0	51	0	0	0	0	11	0	1	0	0	0	0	100
Q	70.00	0	0	70	0	0	0	0	0	0	9	0	15	0	100
R	88.00	3	0	0	88	0	2	0	0	0	0	0	0	0	100
S	97.00	0	0	0	0	97	0	0	0	0	0	0	0	[illegible]	100
T	95.00	2	0	0	3	0	95	0	0	0	0	0	0	0	100
U	83.00	0	2	0	0	0	0	[illegible]	0	0	0	0	0	0	100
V	78.00	0	0	0	0	0	0	0	78	0	0	19	0	0	100
W	61.00	0	0	0	0	0	0	0	0	61	0	0	0	0	100
X	62.00	0	0	5	0	0	0	0	0	3	62	0	0	0	100
Y	86.00	0	0	0	0	0	0	0	13	0	0	86	0	0	100
Z	22.00	0	0	27	0	0	0	0	0	0	35	0	22	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	70.78	156	89	123	107	100	105	104	103	103	142	107	54	100	2700

Table B.7: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	27.00	27	0	0	0	9	0	0	0	0	0	0	0	0	5
B	0.00	0	0	0	0	0	3	27	0	0	0	3	1	0	0
C	26.00	0	0	26	0	0	0	2	0	0	0	0	41	6	0
D	0.00	0	0	12	0	0	1	1	0	0	0	0	49	17	0
E	91.00	1	0	0	0	91	0	0	0	0	0	0	0	0	0
F	11.00	0	0	0	0	0	11	24	0	0	0	0	5	6	0
G	35.00	0	0	0	0	0	14	35	0	0	0	0	0	1	0
H	0.00	0	0	13	0	0	4	2	0	0	0	0	47	6	0
I	0.00	17	0	2	0	2	0	0	0	0	0	0	1	0	5
J	62.00	0	0	0	0	0	0	0	0	0	62	9	0	0	0
K	20.00	0	0	0	0	0	0	1	0	0	8	20	0	0	0
L	52.00	0	0	20	0	0	1	1	0	0	0	0	52	14	0
M	23.00	0	0	4	0	0	2	4	0	0	0	0	37	23	0
N	7.00	19	0	1	0	2	0	0	0	0	0	0	0	0	7
O	22.00	19	0	1	0	3	0	0	0	0	0	0	0	0	9
P	0.00	0	0	1	0	0	7	12	0	0	0	0	23	8	0
Q	0.00	0	0	0	0	0	0	0	0	0	52	13	0	0	0
R	47.00	4	0	7	0	1	0	0	0	0	0	0	1	0	2
S	19.00	12	0	6	0	0	0	0	0	0	0	0	0	0	7
T	46.00	16	0	0	0	12	0	0	0	0	0	0	0	0	4
U	43.00	0	0	1	0	0	8	11	0	0	0	0	19	15	0
V	50.00	0	0	0	0	0	1	8	0	0	0	13	0	0	0
W	20.00	0	0	0	0	0	7	15	0	0	0	4	0	1	0
X	27.00	0	0	0	0	0	0	0	0	0	28	22	0	0	0
Y	24.00	0	0	0	0	0	8	28	0	0	0	0	0	0	0
Z	0.00	0	0	0	0	0	0	1	0	0	77	4	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	27.85	115	0	94	0	120	67	172	0	0	227	88	276	97	39

Table B.7 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	27.00	15	0	0	6	7	31	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	8	22	15	0	21	0	0	100
C	26.00	0	0	0	12	0	0	13	0	0	0	0	0	0	100
D	0.00	0	0	0	7	1	0	12	0	0	0	0	0	0	100
E	91.00	0	0	0	0	0	6	0	0	0	0	0	0	2	100
F	11.00	0	0	0	0	0	0	35	1	8	0	10	0	0	100
G	35.00	0	0	0	0	0	0	18	10	13	0	9	0	0	100
H	0.00	0	0	0	2	0	0	26	0	0	0	0	0	0	100
I	0.00	19	0	0	29	15	10	0	0	0	0	0	0	0	100
J	62.00	0	0	0	0	0	0	0	2	0	27	0	0	0	100
K	20.00	0	0	0	0	0	0	1	38	7	22	3	0	0	100
L	52.00	0	0	0	4	0	0	8	0	0	0	0	0	0	100
M	23.00	0	0	0	2	0	0	27	0	0	0	1	0	0	100
N	7.00	19	0	0	20	18	14	0	0	0	0	0	0	0	100
O	22.00	22	0	0	17	12	17	0	0	0	0	0	0	0	100
P	0.00	0	0	0	1	0	0	46	1	0	0	1	0	0	100
Q	0.00	0	0	0	0	0	0	0	4	0	31	0	0	0	100
R	47.00	17	0	0	47	18	3	0	0	0	0	0	0	0	100
S	19.00	20	0	0	27	19	9	0	0	0	0	0	0	0	100
T	46.00	13	0	0	3	6	46	0	0	0	0	0	0	0	100
U	43.00	0	0	0	0	0	0	43	0	1	0	2	0	0	100
V	50.00	0	0	0	0	0	0	1	50	10	0	17	0	0	100
W	20.00	0	0	0	0	0	0	3	22	20	0	28	0	0	100
X	27.00	0	0	0	0	0	0	0	22	0	27	1	0	0	100
Y	24.00	0	0	0	0	0	0	7	17	16	0	24	0	0	100
Z	0.00	0	0	0	0	0	0	0	1	0	17	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	27.85	125	0	0	177	96	136	248	190	90	124	117	0	102	2700

Table B.8: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	14.00	14	0	2	0	3	0	0	0	0	0	0	0	0	5
B	0.00	0	0	0	0	0	2	26	0	0	0	0	0	0	0
C	25.00	0	0	25	0	0	1	0	0	0	0	0	37	14	0
D	0.00	0	0	14	0	0	1	1	0	0	0	0	45	20	0
E	87.00	0	0	0	0	87	0	0	0	0	0	0	0	0	0
F	11.00	0	0	0	0	0	11	34	0	0	0	1	6	2	0
G	33.00	0	0	0	0	0	8	33	0	0	0	1	1	2	0
H	0.00	0	0	12	0	0	1	5	0	0	0	0	41	12	0
I	0.00	12	0	4	0	2	0	0	0	0	0	0	0	0	8
J	66.00	0	0	0	0	0	0	0	0	0	66	8	0	0	0
K	24.00	0	0	0	0	0	0	1	0	0	6	24	0	0	0
L	42.00	0	0	23	0	0	0	3	0	0	0	0	42	17	0
M	11.00	0	0	9	0	0	5	5	0	0	0	0	28	11	0
N	7.00	20	0	1	0	2	0	0	0	0	0	0	0	0	7
O	32.00	15	0	0	0	3	0	0	0	0	0	0	0	0	3
P	0.00	0	0	0	0	0	11	13	0	0	0	0	19	20	0
Q	0.00	0	0	0	0	0	0	0	0	0	62	11	0	0	0
R	40.00	6	0	10	0	0	0	0	0	0	0	0	0	0	4
S	13.00	12	0	4	0	1	0	0	0	0	0	0	0	0	14
T	43.00	18	0	1	0	22	0	0	0	0	0	0	0	0	1
U	46.00	0	0	1	0	0	8	10	0	0	0	0	10	20	0
V	41.00	0	0	0	0	0	1	5	0	0	1	7	0	0	0
W	15.00	0	0	0	0	0	5	32	0	0	0	3	0	0	0
X	22.00	0	0	0	0	0	0	0	0	0	21	33	0	0	0
Y	27.00	0	0	0	0	0	3	24	0	0	0	2	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	79	3	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	25.89	97	0	106	0	120	57	192	0	0	235	93	229	118	42

Table B.8 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	14.00	21	0	0	8	7	40	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	5	19	26	0	22	0	0	100
C	25.00	0	0	0	19	0	0	4	0	0	0	0	0	0	100
D	0.00	0	0	0	5	0	0	14	0	0	0	0	0	0	100
E	87.00	0	0	0	0	0	13	0	0	0	0	0	0	0	100
F	11.00	0	0	0	0	0	0	35	1	1	1	8	0	0	100
G	33.00	0	0	0	0	0	0	20	7	5	0	23	0	0	100
H	0.00	0	0	0	1	0	0	26	0	2	0	0	0	0	100
I	0.00	23	0	0	20	21	10	0	0	0	0	0	0	0	100
J	66.00	0	0	0	0	0	0	0	0	0	26	0	0	0	100
K	24.00	0	0	0	0	0	0	0	47	5	15	2	0	0	100
L	42.00	0	0	0	6	1	0	8	0	0	0	0	0	0	100
M	11.00	0	0	0	1	0	0	40	0	1	0	0	0	0	100
N	7.00	21	0	0	17	17	15	0	0	0	0	0	0	0	100
O	32.00	32	0	0	19	9	19	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	33	0	1	0	3	0	0	100
Q	0.00	0	0	0	0	0	0	0	9	0	18	0	0	0	100
R	40.00	17	0	0	40	21	2	0	0	0	0	0	0	0	100
S	13.00	25	0	0	23	13	8	0	0	0	0	0	0	0	100
T	43.00	9	0	0	5	1	43	0	0	0	0	0	0	0	100
U	46.00	0	0	0	0	0	0	46	0	0	0	5	0	0	100
V	41.00	0	0	0	0	0	0	2	41	22	2	19	0	0	100
W	15.00	0	0	0	0	0	0	6	25	15	1	13	0	0	100
X	22.00	0	0	0	0	0	0	0	20	1	22	3	0	0	100
Y	27.00	0	0	0	0	0	0	8	23	13	0	27	0	0	100
Z	0.00	0	0	0	0	0	0	0	1	0	17	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	25.89	148	0	0	164	90	150	247	193	92	102	125	0	100	2700

Table B.9: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	20.00	20	0	0	0	10	0	0	0	0	0	0	0	0	4
B	0.00	0	0	0	0	0	16	27	0	0	0	1	1	1	0
C	24.00	0	0	24	0	0	1	1	0	0	0	0	44	10	0
D	0.00	0	0	10	0	0	0	1	0	0	0	0	56	18	0
E	92.00	1	0	0	0	92	0	0	0	0	0	0	0	0	0
F	19.00	0	0	0	0	0	19	16	0	0	0	0	7	14	0
G	31.00	0	0	0	0	0	13	31	0	0	0	0	0	5	0
H	0.00	0	0	13	0	0	1	1	0	0	0	6	49	13	0
I	0.00	15	0	2	0	2	0	0	0	0	0	0	1	0	8
J	0.00	0	0	0	0	0	0	0	0	0	0	58	0	0	0
K	45.00	0	0	0	0	0	0	1	0	0	0	45	0	1	0
L	59.00	0	0	19	0	0	0	1	0	0	0	0	59	14	0
M	23.00	0	0	4	0	0	3	2	0	0	0	0	47	23	0
N	7.00	9	0	1	0	2	0	0	0	0	0	0	0	0	7
O	29.00	13	0	1	0	3	0	0	0	0	0	0	0	0	6
P	0.00	0	0	1	0	0	7	6	0	0	0	0	28	20	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	62	0	0	0
R	44.00	2	0	5	0	1	0	0	0	0	0	0	1	0	7
S	11.00	17	0	5	0	1	0	0	0	0	0	0	0	0	9
T	50.00	17	0	0	0	13	0	0	0	0	0	0	0	0	5
U	31.00	0	0	1	0	0	6	7	0	0	0	0	22	32	0
V	49.00	0	0	0	0	0	3	20	0	0	0	4	0	0	0
W	22.00	0	0	0	0	0	6	30	0	0	0	2	1	0	0
X	1.00	0	0	0	0	0	0	1	0	0	0	57	0	0	0
Y	13.00	0	0	0	0	0	12	35	0	0	0	0	0	0	0
Z	55.00	0	0	0	0	0	0	1	0	0	0	36	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	26.85	94	0	86	0	124	87	181	0	0	0	265	316	151	46

Table B.9 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	20.00	16	0	0	4	4	42	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	10	18	12	0	14	0	0	100
C	24.00	0	0	0	14	0	0	6	0	0	0	0	0	0	100
D	0.00	0	0	0	8	2	0	5	0	0	0	0	0	0	100
E	92.00	0	0	0	0	0	5	0	0	0	0	0	0	2	100
F	19.00	0	0	0	0	0	0	35	1	3	0	5	0	0	100
G	31.00	0	0	0	0	0	0	27	4	13	0	7	0	0	100
H	0.00	0	0	0	3	0	0	20	0	0	0	0	0	0	100
I	0.00	20	0	0	25	10	17	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	2	0	5	0	35	0	100
K	45.00	0	0	0	0	0	0	0	34	10	0	8	1	0	100
L	59.00	0	0	0	5	0	0	2	0	0	0	0	0	0	100
M	23.00	0	0	0	3	0	0	18	0	0	0	0	0	0	100
N	7.00	26	0	0	15	11	29	0	0	0	0	0	0	0	100
O	29.00	29	0	0	14	7	27	0	0	0	0	0	0	0	100
P	0.00	0	0	0	1	0	0	36	1	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	6	0	5	0	27	0	100
R	44.00	21	0	0	44	13	6	0	0	0	0	0	0	0	100
S	11.00	20	0	0	25	11	12	0	0	0	0	0	0	0	100
T	50.00	11	0	0	1	3	50	0	0	0	0	0	0	0	100
U	31.00	0	0	0	0	0	0	31	0	0	0	1	0	0	100
V	49.00	0	0	0	0	0	0	2	49	13	0	9	0	0	100
W	22.00	0	0	0	0	0	0	10	11	22	0	18	0	0	100
X	1.00	0	0	0	0	0	0	0	26	1	1	0	14	0	100
Y	13.00	0	0	0	0	0	0	15	10	15	0	13	0	0	100
Z	55.00	0	0	0	0	0	0	0	1	0	7	0	55	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	26.85	143	0	0	162	61	188	217	163	89	18	75	132	102	2700

Table B.10: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	15.00	15	0	1	0	4	0	0	0	0	0	0	0	0	4
B	0.00	0	0	0	0	0	8	33	0	0	0	0	0	1	0
C	26.00	0	0	26	0	0	0	0	0	0	0	0	44	5	0
D	0.00	0	0	14	0	0	0	1	0	0	0	0	57	16	0
E	88.00	0	0	0	0	88	0	0	0	0	0	0	0	0	0
F	18.00	0	0	0	0	0	18	24	0	0	0	2	6	11	0
G	37.00	0	0	0	0	0	17	37	0	0	0	1	1	5	0
H	0.00	0	0	13	0	0	2	3	0	0	0	0	44	20	0
I	0.00	14	0	3	0	3	0	0	0	0	0	0	0	0	10
J	0.00	0	0	0	0	0	0	0	0	0	0	54	0	0	0
K	37.00	0	0	0	0	0	0	2	0	0	0	37	0	0	0
L	51.00	0	0	23	0	0	3	0	0	0	0	0	51	13	0
M	20.00	0	0	10	0	0	3	2	0	0	0	0	33	20	0
N	12.00	20	0	1	0	2	0	0	0	0	0	0	0	0	12
O	24.00	17	0	0	0	4	0	0	0	0	0	0	0	0	5
P	0.00	0	0	0	0	0	11	4	0	0	0	0	24	29	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	54	0	0	0
R	38.00	7	0	7	0	0	0	0	0	0	0	0	0	0	15
S	14.00	9	0	3	0	2	0	0	0	0	0	0	0	0	6
T	50.00	11	0	1	0	24	0	0	0	0	0	0	0	0	0
U	32.00	0	0	1	0	0	4	11	0	0	0	0	19	33	0
V	22.00	0	0	0	0	0	2	15	0	0	0	7	0	1	0
W	16.00	0	0	0	0	0	20	22	0	0	0	2	0	0	0
X	3.00	0	0	0	0	0	0	2	0	0	0	53	0	0	0
Y	15.00	0	0	0	0	0	10	34	0	0	0	0	0	1	0
Z	44.00	0	0	0	0	0	0	0	0	0	0	45	0	0	0
@	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	24.48	93	0	103	0	128	98	190	0	0	0	255	279	155	52

Table B.10 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	15.00	17	0	0	6	5	48	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	6	10	20	0	22	0	0	100
C	26.00	0	0	0	19	1	0	5	0	0	0	0	0	0	100
D	0.00	0	0	0	5	0	0	7	0	0	0	0	0	0	100
E	88.00	0	0	0	0	0	12	0	0	0	0	0	0	0	100
F	18.00	0	0	0	0	0	0	36	1	0	0	2	0	0	100
G	37.00	0	0	0	0	0	0	25	2	6	0	6	0	0	100
H	0.00	0	0	0	3	0	0	13	0	1	0	1	0	0	100
I	0.00	24	0	0	18	13	15	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	0	0	12	0	34	0	100
K	37.00	0	0	0	0	0	0	0	40	17	0	3	1	0	100
L	51.00	0	0	0	6	1	0	3	0	0	0	0	0	0	100
M	20.00	0	0	0	1	0	0	30	0	0	0	1	0	0	100
N	12.00	19	0	0	13	9	24	0	0	0	0	0	0	0	100
O	24.00	24	0	0	15	8	27	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	30	0	1	0	1	0	0	100
Q	0.00	0	0	0	0	0	0	0	6	5	9	0	26	0	100
R	38.00	19	0	0	38	10	4	0	0	0	0	0	0	0	100
S	14.00	36	0	0	16	14	[illegible]	0	0	0	0	0	0	0	100
T	50.00	9	0	0	5	0	50	0	0	0	0	0	0	0	100
U	32.00	0	0	0	0	0	0	32	0	0	0	0	0	0	100
V	22.00	0	0	0	0	0	0	2	22	32	0	18	1	0	100
W	16.00	0	0	0	0	0	0	11	18	16	0	11	0	0	100
X	3.00	0	0	0	0	0	0	0	30	1	3	2	9	0	100
Y	15.00	0	0	0	0	0	0	10	13	17	0	15	0	0	100
Z	44.00	0	0	0	0	0	0	0	1	0	10	0	44	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	24.48	148	0	0	145	61	194	210	143	116	34	82	115	99	2700

Table B.11: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 1, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	65.00	65	0	0	0	1	0	0	0	11	0	0	0	0	0
B	30.00	0	30	0	0	0	21	13	2	0	0	0	0	4	0
C	48.00	0	0	48	3	0	0	0	1	0	0	0	21	16	2
D	82.00	0	0	1	82	0	1	0	5	0	0	0	5	5	0
E	97.00	1	0	0	0	97	0	0	0	1	0	0	0	0	0
F	36.00	0	5	7	1	0	36	12	15	0	0	1	0	12	0
G	51.00	0	3	1	3	0	14	51	8	0	0	0	2	6	0
H	89.00	0	0	2	2	0	3	3	89	0	0	0	0	1	0
I	85.00	9	0	0	0	1	0	0	0	85	0	0	0	0	0
J	6.00	0	0	0	0	0	0	0	0	0	6	19	0	0	0
K	33.00	0	10	0	0	0	3	7	0	0	1	33	0	0	0
L	80.00	0	0	3	2	0	0	0	1	1	0	0	80	5	0
M	41.00	0	0	13	4	0	5	1	1	0	0	0	25	41	1
N	95.00	0	0	2	0	0	0	0	0	0	0	0	0	0	95
O	76.00	10	0	0	0	1	0	0	0	2	0	0	1	0	0
P	14.00	0	0	10	3	0	3	3	2	0	0	0	19	30	0
Q	31.00	0	0	0	0	0	0	0	0	0	5	15	0	0	0
R	80.00	0	0	0	1	0	0	0	0	0	0	0	3	0	6
S	87.00	1	0	0	2	2	0	0	0	1	0	0	1	0	0
T	85.00	4	0	0	0	0	0	0	0	0	0	0	0	0	1
U	68.00	0	1	0	1	0	1	0	0	0	0	0	19	7	0
V	60.00	0	0	0	1	0	2	5	0	0	0	3	0	0	0
W	49.00	0	14	0	0	0	17	7	3	0	0	1	0	0	0
X	38.00	0	5	0	0	0	0	1	0	0	4	20	0	0	0
Y	82.00	0	0	0	3	0	0	2	0	0	0	1	0	0	0
Z	78.00	0	0	0	0	0	0	1	0	0	2	8	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	62.44	90	68	87	108	102	106	106	127	101	18	101	176	127	105

Table B.11 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	65.00	17	0	0	2	0	4	6	0	0	0	0	0	0	100
B	30.00	0	2	0	0	0	0	0	5	23	0	0	0	0	100
C	48.00	0	0	0	7	0	0	2	0	0	0	0	0	0	100
D	82.00	0	0	0	0	0	0	0	0	0	0	1	0	0	100
E	97.00	0	0	0	0	1	0	0	0	0	0	0	0	0	100
F	36.00	0	4	0	0	0	0	2	0	5	0	0	0	0	100
G	51.00	0	4	0	0	0	0	0	1	6	0	1	0	0	100
H	89.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
I	85.00	1	0	0	2	1	1	0	0	0	0	0	0	0	100
J	6.00	0	0	3	0	0	0	0	0	0	16	0	56	0	100
K	33.00	0	0	2	0	0	0	0	24	11	2	2	5	0	100
L	80.00	1	2	0	2	0	0	3	0	0	0	0	0	0	100
M	41.00	0	4	0	0	0	0	5	0	0	0	0	0	0	100
N	95.00	0	0	0	0	0	3	0	0	0	0	0	0	0	100
O	76.00	76	0	0	7	0	3	0	0	0	0	0	0	0	100
P	14.00	0	14	0	0	0	0	14	1	1	0	0	0	0	100
Q	31.00	0	0	31	0	0	0	0	0	1	6	0	42	0	100
R	80.00	2	0	0	80	1	7	0	0	0	0	3	0	0	100
S	87.00	0	0	0	5	87	1	0	0	0	0	0	0	0	100
T	85.00	7	0	0	3	0	85	0	0	0	0	0	0	0	100
U	68.00	0	2	0	0	0	0	68	0	0	0	1	0	0	100
V	60.00	0	0	0	0	0	0	0	60	4	1	24	0	0	100
W	49.00	0	1	0	0	0	0	0	6	49	0	2	0	0	100
X	38.00	0	0	1	0	0	0	0	5	7	38	0	19	0	100
Y	82.00	0	0	0	0	0	0	1	10	1	0	82	0	0	100
Z	78.00	0	0	1	0	0	0	0	1	0	9	0	78	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	62.44	104	33	38	108	90	104	95	113	108	72	113	200	100	2700

Table B.12: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 1, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	58.00	58	0	1	0	0	0	0	0	13	0	0	0	0	0
B	27.00	0	27	0	0	0	23	12	0	0	0	0	[illegible]	1	0
C	44.00	0	0	44	4	0	0	0	2	0	0	0	24	12	3
D	72.00	0	0	8	72	0	1	0	5	0	0	0	9	3	0
E	96.00	0	0	0	0	96	0	0	0	2	0	0	0	0	0
F	41.00	0	5	4	1	0	41	11	12	0	0	2	0	12	0
G	39.00	0	2	3	1	0	20	39	11	0	0	0	0	11	0
H	85.00	0	4	1	3	0	2	0	85	0	0	0	0	3	0
I	85.00	10	0	0	0	0	0	0	0	85	0	0	0	0	0
J	7.00	0	1	0	0	0	0	0	0	0	7	26	0	0	0
K	29.00	0	4	0	0	0	1	6	0	0	1	29	0	0	0
L	80.00	0	0	1	2	0	0	0	0	1	0	0	80	2	0
M	30.00	0	2	19	5	0	4	2	2	0	0	0	19	30	0
N	92.00	0	0	1	0	0	0	0	0	0	0	0	0	0	92
O	75.00	7	0	0	0	3	0	0	0	7	0	0	0	0	0
P	13.00	0	2	12	2	0	3	5	2	0	0	0	20	30	0
Q	19.00	0	0	0	0	0	0	1	0	0	2	17	0	0	0
R	72.00	0	0	2	2	0	0	0	0	1	0	0	1	0	5
S	90.00	0	0	0	2	2	0	0	0	1	0	0	0	0	0
T	84.00	6	0	1	1	0	0	0	0	0	0	0	0	0	0
U	68.00	0	0	0	0	0	1	2	1	0	0	0	19	4	0
V	45.00	0	0	0	0	0	0	7	0	0	0	2	0	0	0
W	42.00	0	11	0	0	0	23	13	0	0	0	3	0	0	0
X	35.00	0	12	0	0	0	0	0	0	0	2	15	0	0	0
Y	66.00	0	1	0	2	0	1	3	0	0	0	1	0	1	0
Z	71.00	0	0	0	0	0	0	0	0	0	3	12	0	0	0
[illegible]	98.00	0	0	0	0	2	0	0	0	0	0	0	0	0	0
TOTAL	57.89	81	71	97	97	103	120	101	120	110	15	107	172	109	100

Table B.12 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	58.00	24	0	0	1	0	3	0	0	0	0	0	0	0	100
B	27.00	0	5	0	0	0	0	1	1	28	0	2	0	0	100
C	44.00	0	2	0	7	0	0	2	0	0	0	0	0	0	100
D	72.00	0	0	0	0	0	0	2	0	0	0	0	0	0	100
E	96.00	2	0	0	0	0	0	0	0	0	0	0	0	0	100
F	41.00	0	8	0	0	0	0	0	0	4	0	0	0	0	100
G	39.00	0	5	0	0	0	0	1	1	5	0	1	0	0	100
H	85.00	0	0	0	0	0	0	0	0	1	0	1	0	0	100
I	85.00	4	0	0	0	0	0	1	0	0	0	0	0	0	100
J	7.00	0	0	4	0	0	0	0	0	0	10	0	52	0	100
K	29.00	0	2	1	0	0	0	0	29	16	5	4	2	0	100
L	80.00	1	1	0	4	1	0	7	0	0	0	0	0	0	100
M	30.00	0	7	0	1	0	0	8	0	0	0	1	0	0	100
N	92.00	0	0	0	5	0	2	0	0	0	0	0	0	0	100
O	75.00	75	0	0	4	1	3	0	0	0	0	0	0	0	100
P	13.00	0	13	0	0	0	0	11	0	0	0	0	0	0	100
Q	19.00	0	0	19	0	0	0	0	3	0	12	0	46	0	100
R	72.00	10	0	0	72	0	7	0	0	0	0	0	0	0	100
S	90.00	0	0	0	1	90	4	0	0	0	0	0	0	0	100
T	84.00	2	0	0	5	1	84	0	0	0	0	0	0	0	100
U	68.00	0	3	0	1	0	0	68	0	1	0	0	0	0	100
V	45.00	0	0	0	0	0	0	0	45	6	1	37	2	0	100
W	42.00	0	3	0	0	0	0	0	3	42	2	0	0	0	100
X	35.00	0	0	0	0	0	0	0	3	12	35	0	21	0	100
Y	66.00	0	0	0	0	0	0	1	23	1	0	66	0	0	100
Z	71.00	0	0	4	0	0	0	0	0	0	10	0	71	0	100
@	98.00	0	0	0	0	0	0	0	0	0	0	0	0	98	100
TOTAL	57.89	118	49	28	101	93	103	102	108	116	75	112	194	98	2700

Table B.13: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	28.00	28	0	0	0	6	0	0	0	10	0	0	0	0	0
B	0.00	0	0	0	0	0	11	10	0	0	0	3	0	0	0
C	37.00	0	0	37	21	0	1	0	7	1	0	0	10	0	0
D	43.00	0	0	19	43	0	3	1	7	0	0	0	12	0	0
E	84.00	0	0	0	0	84	0	0	0	0	0	0	0	0	0
F	31.00	0	0	0	4	0	31	10	2	0	0	0	0	0	0
G	13.00	0	0	1	1	0	32	13	0	0	0	2	0	0	0
H	8.00	0	0	12	27	0	6	2	8	0	0	0	7	0	0
I	20.00	10	0	2	0	1	0	0	0	20	0	0	0	0	0
J	27.00	0	0	0	0	0	0	0	0	0	27	17	0	0	0
K	54.00	0	0	0	0	0	2	1	0	0	2	54	0	0	0
L	13.00	0	0	22	37	0	1	0	5	0	0	0	13	0	0
M	0.00	0	0	6	19	0	11	3	13	0	0	1	5	0	0
N	0.00	12	0	2	0	3	0	0	0	25	0	0	0	0	0
O	26.00	24	0	2	0	1	0	0	0	19	0	0	0	0	0
P	0.00	0	0	7	13	0	14	4	8	0	0	0	1	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	28	8	0	0	0
R	51.00	5	0	5	1	1	0	0	0	24	0	0	0	0	0
S	0.00	14	0	3	1	1	0	0	0	10	0	0	0	0	0
T	40.00	28	0	0	0	13	0	0	0	5	0	0	0	0	0
U	49.00	0	0	2	17	0	17	1	7	0	0	0	2	0	0
V	49.00	0	0	0	0	0	5	3	0	0	1	12	0	0	0
W	19.00	0	0	0	0	0	17	8	0	0	0	0	0	0	0
X	23.00	0	0	0	0	0	0	0	0	0	16	40	0	0	0
Y	19.00	0	0	0	0	0	25	9	0	0	0	3	0	0	0
Z	59.00	0	0	0	0	0	0	0	0	0	20	2	0	0	0
@	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	29.33	121	0	120	184	111	176	65	57	114	94	142	50	0	0

Table B.13 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	28.00	21	0	0	7	0	28	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	6	31	19	1	19	0	0	100
C	37.00	0	0	0	11	0	0	12	0	0	0	0	0	0	100
D	43.00	0	0	0	4	0	0	11	0	0	0	0	0	0	100
E	84.00	0	0	0	0	0	14	0	0	0	0	0	0	2	100
F	31.00	0	0	0	0	0	0	28	6	8	0	11	0	0	100
G	13.00	0	0	0	0	0	0	13	10	14	0	14	0	0	100
H	8.00	0	0	0	2	0	0	36	0	0	0	0	0	0	100
I	20.00	28	0	0	28	0	11	0	0	0	0	0	0	0	100
J	27.00	0	0	0	0	0	0	0	0	0	23	0	33	0	100
K	54.00	0	0	0	0	0	0	0	24	3	9	1	4	0	100
L	13.00	0	0	0	9	0	0	13	0	0	0	0	0	0	100
M	0.00	0	0	0	1	0	0	39	0	0	0	2	0	0	100
N	0.00	16	0	0	35	0	7	0	0	0	0	0	0	0	100
O	26.00	26	0	0	16	0	12	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	48	2	1	0	2	0	0	100
Q	0.00	0	0	0	0	0	0	0	3	0	24	0	37	0	100
R	51.00	11	0	0	51	0	2	0	0	0	0	0	0	0	100
S	0.00	24	0	0	39	0	8	0	0	0	0	0	0	0	100
T	40.00	9	0	0	5	0	40	0	0	0	0	0	0	0	100
U	49.00	0	0	0	0	0	0	49	0	2	0	3	0	0	100
V	49.00	0	0	0	0	0	0	0	49	20	0	10	0	0	100
W	19.00	0	0	0	0	0	0	4	37	19	0	15	0	0	100
X	23.00	0	0	0	0	0	0	0	10	0	23	0	11	0	100
Y	19.00	0	0	0	0	0	0	5	20	19	0	19	0	0	100
Z	59.00	0	0	0	0	0	0	0	1	0	18	0	59	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	29.33	135	0	0	208	0	122	264	193	105	98	96	144	101	2700

Table B.14: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.00	24	0	0	0	4	0	0	0	12	0	0	0	0	0
B	0.00	0	0	0	1	0	21	9	0	0	0	5	0	0	0
C	23.00	0	0	23	31	0	0	0	4	1	0	0	15	0	0
D	46.00	0	0	17	46	0	1	1	5	0	0	0	12	0	0
E	91.00	0	0	0	0	91	0	0	0	0	0	0	0	0	0
F	42.00	0	0	0	1	0	42	7	1	0	0	0	0	0	0
G	17.00	0	0	1	1	0	25	17	0	0	0	0	0	0	0
H	5.00	0	0	15	30	0	10	1	5	0	0	1	7	0	0
I	26.00	17	0	1	1	1	0	0	0	26	0	0	0	0	0
J	37.00	0	0	0	0	0	0	0	0	0	37	6	0	0	0
K	47.00	0	0	0	0	0	0	0	0	0	3	47	0	0	0
L	13.00	0	0	24	36	0	0	0	6	0	0	0	13	0	0
M	0.00	0	0	7	16	0	17	0	7	0	0	0	5	0	0
N	0.00	19	0	3	0	2	0	0	0	21	0	0	0	0	0
O	26.00	20	0	0	0	2	0	0	0	18	0	0	0	0	0
P	0.00	0	0	3	14	0	25	4	9	0	0	2	2	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	28	8	0	0	0
R	59.00	2	0	8	1	0	0	0	0	20	0	0	0	0	0
S	0.00	17	0	0	0	2	0	0	0	18	0	0	1	0	0
T	47.00	16	0	0	0	19	0	0	0	4	0	0	0	0	0
U	45.00	0	0	4	16	0	17	6	7	0	0	2	1	0	0
V	46.00	0	0	0	0	0	6	4	0	0	0	23	0	0	0
W	22.00	0	0	0	0	0	17	9	0	0	0	2	0	0	0
X	29.00	0	0	0	0	0	0	0	0	0	13	31	0	0	0
Y	9.00	0	0	0	0	0	18	13	0	0	0	3	0	0	0
Z	44.00	0	0	0	0	0	0	0	0	0	29	6	0	0	0
@	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	29.52	115	0	106	194	122	193	71	44	120	110	136	56	0	0

Table B.14 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	24.00	20	0	0	5	0	35	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	3	26	22	1	12	0	0	100
C	23.00	0	0	0	11	0	0	15	0	0	0	0	0	0	100
D	46.00	0	0	0	5	0	0	12	0	1	0	0	0	0	100
E	91.00	0	0	0	0	0	9	0	0	0	0	0	0	0	100
F	42.00	0	0	0	0	0	0	32	2	7	0	8	0	0	100
G	17.00	0	0	0	0	0	0	19	9	16	0	12	0	0	100
H	5.00	0	0	0	5	0	0	25	0	0	0	1	0	0	100
I	26.00	22	0	0	27	0	5	0	0	0	0	0	0	0	100
J	37.00	0	0	0	0	0	0	0	2	0	19	0	36	0	100
K	47.00	0	0	0	0	0	0	0	33	2	11	3	1	0	100
L	13.00	1	0	0	6	0	0	14	0	0	0	0	0	0	100
M	0.00	0	0	0	2	0	0	45	0	1	0	0	0	0	100
N	0.00	18	0	0	27	0	10	0	0	0	0	0	0	0	100
O	26.00	26	0	0	17	0	17	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	39	1	0	0	1	0	0	100
Q	0.00	0	0	0	0	0	0	0	4	0	23	0	37	0	100
R	59.00	9	0	0	59	0	1	0	0	0	0	0	0	0	100
S	0.00	19	0	0	38	0	5	0	0	0	0	0	0	0	100
T	47.00	11	0	0	3	0	47	0	0	0	0	0	0	0	100
U	45.00	0	0	0	0	0	0	45	1	1	0	0	0	0	100
V	46.00	0	0	0	0	0	0	0	46	16	2	8	1	0	100
W	22.00	0	0	0	0	0	0	10	27	22	0	13	0	0	100
X	29.00	0	0	0	0	0	0	0	12	2	29	0	13	0	100
Y	9.00	0	0	0	0	0	0	3	36	18	0	9	0	0	100
Z	44.00	0	0	0	0	0	0	0	0	0	21	0	44	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	29.52	126	0	0	205	0	129	262	199	108	106	67	132	99	2700

Table B.15: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	L	E	F	G	H	I	J	K	L	M	N
A	29.00	29	0	0	0	9	0	0	0	13	0	0	0	0	0
B	0.00	0	0	0	0	0	23	21	1	0	0	2	0	0	0
C	32.00	0	0	32	12	0	1	0	9	1	0	0	23	0	0
D	25.00	0	0	18	25	0	1	0	8	0	0	0	35	0	0
E	84.00	0	0	0	0	84	0	0	0	0	0	0	0	0	0
F	35.00	0	0	0	4	0	35	8	4	0	0	0	2	0	0
G	18.00	0	0	1	0	0	38	18	4	0	0	0	1	0	0
H	21.00	0	0	12	19	0	5	0	21	1	0	0	21	0	0
I	15.00	17	0	2	0	1	0	0	0	15	0	0	0	0	2
J	0.00	0	0	0	0	0	0	0	0	0	0	56	0	0	0
K	37.00	0	0	0	0	0	3	1	0	0	0	37	0	0	0
L	36.00	0	0	20	19	0	0	0	8	0	0	0	36	0	0
M	0.00	0	0	6	17	0	8	2	23	1	0	1	16	0	0
N	3.00	17	0	2	0	3	0	0	0	24	0	0	0	0	3
O	19.00	29	0	2	0	1	0	0	0	15	0	0	0	0	2
P	0.00	0	0	7	14	0	9	2	22	0	0	0	5	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	54	0	0	0
R	44.00	6	0	4	1	1	0	0	0	24	0	0	0	0	2
S	0.00	26	0	2	0	1	0	0	0	13	0	0	1	0	4
T	49.00	25	0	0	0	14	0	0	0	4	0	0	0	0	1
U	33.00	0	0	2	16	0	10	1	27	0	0	0	9	0	0
V	43.00	0	0	0	0	0	10	12	0	0	0	2	0	0	0
W	34.00	0	0	0	0	0	20	20	0	0	0	0	0	0	0
X	1.00	0	0	0	0	0	0	0	0	0	0	62	0	0	0
Y	0.00	0	0	0	0	0	32	22	0	0	0	1	0	0	0
Z	57.00	0	0	0	0	0	0	0	0	0	0	36	0	0	0
@	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	26.44	149	0	110	127	115	195	107	127	111	0	251	149	0	14

Table B.15 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	29.00	13	0	0	3	0	33	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	8	17	28	0	0	0	0	100
C	32.00	0	0	0	16	0	0	6	0	0	0	0	0	0	100
D	25.00	0	0	0	5	0	0	8	0	0	0	0	0	0	100
E	84.00	0	0	0	0	0	14	0	0	0	0	0	0	2	100
F	35.00	0	0	0	0	0	0	34	1	12	0	0	0	0	100
G	18.00	0	0	0	0	0	0	19	8	11	0	0	0	0	100
H	21.00	0	0	0	1	0	0	20	0	0	0	0	0	0	100
I	15.00	25	0	0	24	0	14	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	5	0	7	0	32	0	100
K	37.00	0	0	0	0	0	0	0	48	7	0	0	4	0	100
L	36.00	0	0	0	10	0	0	7	0	0	0	0	0	0	100
M	0.00	0	0	0	0	0	0	26	0	0	0	0	0	0	100
N	3.00	18	0	0	24	0	9	0	0	0	0	0	0	0	100
O	19.00	19	0	0	11	0	21	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	38	2	1	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	4	1	5	0	36	0	100
R	44.00	13	0	0	44	0	5	0	0	0	0	0	0	0	100
S	0.00	15	0	0	29	0	9	0	0	0	0	0	0	0	100
T	49.00	4	0	0	3	0	49	0	0	0	0	0	0	0	100
U	33.00	0	0	0	0	0	0	33	0	2	0	0	0	0	100
V	43.00	0	0	0	0	0	0	0	43	33	0	0	0	0	100
W	34.00	0	0	0	0	0	0	11	15	34	0	0	0	0	100
X	1.00	0	0	0	0	0	0	0	21	5	1	0	11	0	100
Y	0.00	0	0	0	0	0	0	9	13	23	0	0	0	0	100
Z	57.00	0	0	0	0	0	0	0	2	0	5	0	57	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	26.44	107	0	0	170	0	154	219	179	157	18	0	140	101	2700

Table B.16: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	24.00	24	0	0	0	4	0	0	0	8	0	0	0	0	0
B	0.00	0	0	0	1	0	28	17	0	0	0	3	0	0	0
C	23.00	0	0	23	13	0	0	0	9	1	0	0	35	0	0
D	24.00	0	0	15	24	0	1	1	9	0	0	0	39	0	0
E	93.00	0	0	0	0	93	0	0	0	0	0	0	0	0	0
F	34.00	0	0	0	0	0	34	12	6	0	0	0	1	0	0
G	14.00	0	0	1	1	0	40	14	4	0	0	0	0	0	0
H	16.00	0	0	12	13	0	9	1	16	0	0	0	27	0	0
I	21.00	23	0	0	0	1	0	0	0	21	0	0	1	0	3
J	0.00	0	0	0	0	0	0	0	0	0	0	53	0	0	0
K	35.00	0	0	0	0	0	0	3	0	0	0	35	0	0	0
L	36.00	0	0	22	18	0	0	0	11	0	0	0	36	0	0
M	0.00	0	0	7	14	0	7	1	23	0	0	0	12	0	0
N	2.00	20	0	2	0	2	0	0	0	21	0	0	0	0	2
O	19.00	22	0	0	0	2	0	0	0	16	0	0	0	0	2
P	0.00	0	0	3	13	0	20	0	20	0	0	1	9	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	53	0	0	0
R	53.00	4	0	8	1	0	0	0	0	17	0	0	0	0	1
S	0.00	22	0	0	0	3	0	0	0	16	0	0	1	0	4
T	50.00	18	0	0	0	21	0	0	0	2	0	0	0	0	0
U	38.00	0	0	4	17	0	13	0	19	0	0	0	5	0	0
V	42.00	0	0	0	0	0	4	13	0	0	0	9	0	0	0
W	24.00	0	0	0	0	0	25	18	0	0	0	0	0	0	0
X	1.00	0	0	0	0	0	0	0	0	0	0	60	0	0	0
Y	0.00	0	0	0	0	0	27	17	1	0	0	0	0	0	0
Z	43.00	0	0	0	0	0	0	0	0	0	0	48	0	0	0
@	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	25.59	133	0	97	115	127	208	97	118	102	0	262	166	0	12

Table B.16 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	24.00	15	0	0	5	0	44	0	0	0	0	0	0	0	100
B	0.00	0	0	0	0	0	0	8	16	27	0	0	0	0	100
C	23.00	0	0	0	11	0	0	8	0	0	0	0	0	0	100
D	24.00	0	0	0	6	0	0	5	0	0	0	0	0	0	100
E	93.00	0	0	0	0	0	7	0	0	0	0	0	0	0	100
F	34.00	0	0	0	0	0	0	44	0	3	0	0	0	0	100
G	14.00	0	0	0	0	0	0	21	2	17	0	0	0	0	100
H	16.00	0	0	0	8	0	0	13	1	0	0	0	0	0	100
I	21.00	22	0	0	19	0	10	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	4	0	9	0	34	0	100
K	35.00	0	0	0	0	0	0	0	45	16	0	0	1	0	100
L	36.00	1	0	0	8	0	0	4	0	0	0	0	0	0	100
M	0.00	0	0	0	2	0	0	34	0	0	0	0	0	0	100
N	2.00	16	0	0	18	0	19	0	0	0	0	0	0	0	100
O	19.00	19	0	0	13	0	26	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	32	1	1	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	3	2	6	0	36	0	100
R	53.00	15	0	0	53	0	1	0	0	0	0	0	0	0	100
S	0.00	18	0	0	28	0	8	0	0	0	0	0	0	0	100
T	50.00	8	0	0	1	0	50	0	0	0	0	0	0	0	100
U	38.00	0	0	0	0	0	0	38	2	2	0	0	0	0	100
V	42.00	0	0	0	0	0	0	0	42	31	0	0	1	0	100
W	24.00	0	0	0	0	0	0	15	18	24	0	0	0	0	100
X	1.00	0	0	0	0	0	0	0	21	5	1	0	13	0	100
Y	0.00	0	0	0	0	0	0	7	21	27	0	0	0	0	100
Z	43.00	0	0	0	0	0	0	0	1	0	8	0	43	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	25.59	114	0	0	172	0	165	229	177	155	24	0	128	99	2700

Table B.17: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 2, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	[illegible]	M	N
A	68.00	68	0	0	0	1	0	0	0	12	0	0	0	0	0
B	15.00	0	15	1	0	0	30	8	0	0	0	2	0	4	0
C	58.00	0	0	58	1	0	3	1	1	0	0	0	23	4	2
D	84.00	0	0	2	84	0	1	0	4	0	0	0	4	1	0
E	99.00	1	0	0	0	99	0	0	0	0	0	0	0	0	0
F	62.00	0	1	2	5	0	62	5	12	0	0	0	0	3	0
G	25.00	0	1	4	2	0	34	25	9	0	0	1	0	1	0
H	88.00	0	1	0	3	0	4	2	88	0	0	0	0	0	0
I	77.00	17	0	0	0	0	0	0	0	77	0	0	1	0	0
J	4.00	0	2	0	0	0	0	0	0	0	4	8	0	0	0
K	41.00	0	0	0	0	0	4	2	0	0	0	41	0	0	0
L	79.00	0	0	[illegible]	5	0	0	0	1	0	0	0	79	1	0
M	19.00	0	0	18	12	0	3	0	6	0	0	1	24	19	0
N	85.00	0	0	4	0	0	0	0	1	0	0	0	0	0	85
O	72.00	11	0	0	0	1	0	0	0	3	0	0	2	0	0
P	21.00	0	0	19	2	0	7	1	2	0	0	0	31	5	0
Q	24.00	0	1	0	0	0	0	0	0	0	1	6	0	0	0
R	80.00	1	0	1	0	0	0	0	0	0	0	0	2	1	6
S	88.00	0	0	0	4	0	0	0	0	0	0	0	1	0	0
T	85.00	3	0	0	1	0	0	0	0	0	0	0	0	0	0
U	62.00	0	0	1	2	0	0	1	1	0	0	0	27	0	0
V	63.00	0	1	0	0	0	3	8	1	0	0	2	0	0	0
W	56.00	0	1	0	0	0	29	3	2	0	0	3	0	4	0
X	40.00	0	2	0	0	0	0	0	0	0	0	16	0	0	0
Y	76.00	0	0	0	3	0	4	1	0	0	0	1	0	0	0
Z	79.00	0	0	0	0	0	0	0	0	0	2	0	0	0	0
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	61.07	101	25	112	124	101	184	57	128	92	7	81	194	43	93

Table B.17 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	68.00	17	0	0	0	1	1	0	0	0	0	0	0	0	100
B	15.00	0	7	0	0	0	0	0	1	31	0	1	0	0	100
C	58.00	0	1	0	5	0	0	1	0	0	0	0	0	0	100
D	84.00	0	2	0	0	1	0	1	0	0	0	0	0	0	100
E	99.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
F	62.00	0	6	0	0	0	0	0	2	2	0	0	0	0	100
G	25.00	0	9	0	0	0	0	4	3	6	0	1	0	0	100
H	88.00	0	1	0	0	0	0	1	0	0	0	0	0	0	100
I	77.00	3	0	0	1	0	1	0	0	0	0	0	0	0	100
J	4.00	0	0	12	0	0	0	0	0	0	19	0	55	0	100
K	41.00	0	2	4	0	0	0	0	26	11	7	1	2	0	100
L	79.00	0	1	0	7	0	0	4	0	0	0	0	0	0	100
M	19.00	0	8	0	0	0	0	7	0	1	0	1	0	0	100
N	85.00	0	0	0	9	0	1	0	0	0	0	0	0	0	100
O	72.00	72	0	0	5	1	5	0	0	0	0	0	0	0	100
P	21.00	0	21	0	0	0	0	9	1	0	0	2	0	0	100
Q	24.00	0	0	24	0	0	0	0	2	1	7	0	58	0	100
R	80.00	3	1	0	80	0	5	0	0	0	0	0	0	0	100
S	88.00	1	0	0	2	88	4	0	0	0	0	0	0	0	100
T	85.00	7	0	0	3	1	85	0	0	0	0	0	0	0	100
U	62.00	0	6	[illegible]	0	0	0	62	0	0	0	0	0	0	100
V	63.00	0	0	0	0	0	0	0	63	3	0	19	0	0	100
W	56.00	0	2	[illegible]	0	0	0	0	0	56	0	0	0	0	100
X	40.00	0	0	[illegible]	0	0	0	0	3	10	40	0	22	0	100
Y	76.00	0	0	0	0	0	0	1	12	2	0	76	0	0	100
Z	79.00	0	0	8	0	0	0	0	0	1	10	0	79	0	100
@	99.00	0	0	0	1	0	0	0	0	0	0	0	0	99	100
TOTAL	61.07	103	67	55	113	92	102	90	113	124	83	101	216	99	2700

Table B.18: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 2, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	62.00	62	0	0	0	0	0	0	0	11	0	0	0	0	1
B	14.00	0	14	0	0	0	24	8	1	0	0	3	0	4	0
C	52.00	0	0	52	7	0	1	0	3	0	0	0	21	4	3
D	81.00	0	0	6	81	0	0	1	2	0	0	0	5	1	0
E	90.00	9	0	0	0	90	0	0	0	0	0	0	0	0	0
F	57.00	0	1	5	4	0	57	4	16	0	0	0	0	1	0
G	25.00	0	0	2	5	0	29	25	11	0	0	0	0	3	0
H	86.00	0	2	2	0	0	7	3	86	0	0	0	0	0	0
I	75.00	18	0	0	0	0	0	0	0	75	0	0	0	0	0
J	7.00	0	2	0	0	0	0	0	0	0	7	6	0	0	0
K	34.00	0	1	0	0	0	3	6	0	0	0	34	0	1	0
L	75.00	0	0	7	4	0	0	0	0	0	0	0	75	0	0
M	10.00	0	0	19	9	0	8	0	3	0	0	0	23	10	0
N	89.00	0	0	6	0	0	0	0	0	0	0	0	0	0	89
O	68.00	16	0	1	1	0	0	0	0	3	0	0	0	0	0
P	29.00	0	0	17	3	0	3	4	3	0	0	1	19	9	0
Q	32.00	0	2	0	0	0	0	0	0	0	1	9	0	0	0
R	68.00	2	0	2	1	0	0	0	0	1	0	0	7	0	4
S	86.00	2	0	0	3	2	0	0	0	0	0	0	1	0	0
T	83.00	9	0	1	1	0	0	0	0	1	0	0	0	0	0
U	63.00	0	0	0	2	0	0	1	0	0	0	1	25	1	0
V	58.00	0	0	0	0	0	1	4	0	0	0	1	0	1	0
W	42.00	0	6	0	0	0	37	4	3	0	0	4	0	2	0
X	33.00	0	3	0	0	0	0	0	0	0	1	21	0	0	0
Y	70.00	0	0	0	1	0	3	4	1	0	0	1	0	0	0
Z	66.00	0	0	0	0	0	0	0	0	0	3	7	0	0	0
Θ	99.00	0	0	0	0	1	0	0	0	0	0	0	0	0	0
TOTAL	57.56	118	31	120	122	93	173	64	129	91	12	88	176	37	97

Table B.18 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	62.00	17	0	0	0	1	8	0	0	0	0	0	0	0	100
B	14.00	0	6	0	0	0	0	1	1	35	0	3	0	0	100
C	52.00	0	5	0	3	0	0	0	0	0	0	1	0	0	100
D	81.00	0	1	0	0	1	0	0	1	0	0	1	0	0	100
E	90.00	0	0	0	0	0	1	0	0	0	0	0	0	0	100
F	57.00	0	9	0	0	0	0	0	1	2	0	0	0	0	100
G	25.00	0	18	0	0	0	0	1	1	4	0	1	0	0	100
H	86.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
I	75.00	5	0	0	0	1	0	1	0	0	0	0	0	0	100
J	7.00	0	0	11	0	0	0	0	1	1	18	0	54	0	100
K	34.00	0	0	0	0	0	0	0	27	14	8	2	4	0	100
L	75.00	0	2	0	5	2	0	5	0	0	0	0	0	0	100
M	10.00	0	14	0	0	0	0	13	0	0	0	1	0	0	100
N	89.00	0	0	0	5	0	0	0	0	0	0	0	0	0	100
O	68.00	68	0	0	3	1	7	0	0	0	0	0	0	0	100
P	29.00	0	29	1	0	0	0	9	0	2	0	0	0	0	100
Q	32.00	0	0	32	0	0	0	0	1	0	8	0	47	0	100
R	68.00	6	1	0	68	3	5	0	0	0	0	0	0	0	100
S	86.00	1	0	0	2	86	3	0	0	0	0	0	0	0	100
T	83.00	3	0	0	2	0	83	0	0	0	0	0	0	0	100
U	63.00	0	6	0	0	0	0	63	1	0	0	0	0	0	100
V	58.00	0	0	3	0	0	0	0	58	0	0	30	2	0	100
W	42.00	0	0	0	0	0	0	0	2	42	0	0	0	0	100
X	33.00	0	0	5	0	0	0	0	5	11	33	0	21	0	100
Y	70.00	0	1	0	0	0	6	0	18	1	0	70	0	0	100
Z	66.00	0	0	11	0	0	6	0	0	0	13	0	66	0	100
@	99.00	0	0	0	0	0	0	0	0	0	0	0	0	99	100
TOTAL	57.56	100	92	63	88	95	107	93	117	112	80	109	194	99	2700

Table B.19: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	14.00	14	0	0	0	9	0	0	0	10	0	0	0	0	4
B	14.00	0	14	0	0	0	8	24	0	0	0	5	0	0	0
C	7.00	0	0	7	40	0	2	1	0	0	0	0	23	0	0
D	51.00	0	0	3	51	0	1	0	2	0	0	0	22	0	0
E	86.00	1	0	0	0	86	0	0	0	0	0	0	0	0	0
F	19.00	0	4	0	9	0	19	30	1	0	0	0	0	0	0
G	20.00	0	12	0	0	0	25	20	0	0	0	1	0	0	0
H	1.00	0	0	5	44	0	7	3	1	0	0	0	14	0	0
I	14.00	13	0	4	0	1	0	0	0	14	0	0	1	0	8
J	67.00	0	0	0	0	0	0	0	0	0	67	16	0	0	0
K	59.00	0	5	0	0	0	0	1	0	0	5	59	0	0	0
L	28.00	0	0	7	41	0	1	0	2	0	0	0	28	0	0
M	0.00	0	0	1	31	0	11	4	4	0	0	0	7	0	0
N	7.00	12	0	1	0	0	0	0	0	16	0	0	2	0	7
O	17.00	19	0	0	0	4	0	0	0	17	0	0	1	0	6
P	0.00	0	0	1	27	0	11	9	1	0	0	0	4	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	66	18	0	0	0
R	55.00	2	0	8	0	0	0	0	0	16	0	0	2	0	6
S	0.00	7	0	0	0	2	0	0	0	17	0	0	3	0	9
T	50.00	17	0	0	0	16	0	0	0	4	0	0	0	0	2
U	42.00	0	1	0	25	0	16	10	1	0	0	0	3	0	0
V	46.00	0	12	0	0	0	2	8	0	0	0	16	0	0	0
W	0.00	0	11	0	1	0	10	29	0	0	1	8	0	0	0
X	14.00	0	2	0	0	0	0	0	0	0	27	48	0	0	0
Y	25.00	0	10	0	0	0	18	27	0	0	0	2	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	89	4	0	0	0
α	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	27.26	85	71	37	269	118	131	166	12	94	255	177	110	0	42

Table B.19 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	14.00	13	0	0	7	0	43	0	0	0	0	0	0	0	100
B	14.00	0	0	0	0	0	0	2	23	0	0	24	0	0	100
C	7.00	1	0	0	13	0	0	13	0	0	0	0	0	0	100
D	51.00	0	0	0	4	0	0	17	0	0	0	0	0	0	100
E	86.00	1	0	0	0	0	10	0	0	0	0	0	0	2	100
F	19.00	0	0	0	0	0	0	31	1	0	0	5	0	0	100
G	20.00	0	0	0	0	0	0	10	16	0	0	16	0	0	100
H	1.00	0	0	0	2	0	0	24	0	0	0	0	0	0	100
I	14.00	13	0	0	34	0	12	0	0	0	0	0	0	0	100
J	67.00	0	0	0	0	0	0	0	1	0	16	0	0	0	100
K	59.00	0	0	0	0	0	0	1	25	0	1	3	0	0	100
L	28.00	1	0	0	9	0	0	11	0	0	0	0	0	0	100
M	0.00	0	0	0	4	0	0	37	1	0	0	0	0	0	100
N	7.00	12	0	0	32	0	18	0	0	0	0	0	0	0	100
O	17.00	17	0	0	18	0	18	0	0	0	0	0	0	0	100
P	0.00	0	0	0	2	0	0	42	3	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	3	0	13	0	0	0	100
R	55.00	6	0	0	55	0	5	0	0	0	0	0	0	0	100
S	0.00	13	0	0	37	0	12	0	0	0	0	0	0	0	100
T	50.00	5	0	0	6	0	50	0	0	0	0	0	0	0	100
U	42.00	0	0	0	0	0	0	42	2	0	0	0	0	0	100
V	46.00	0	0	0	0	0	0	1	46	0	2	13	0	0	100
W	0.00	0	0	0	0	0	0	3	18	0	1	18	0	0	100
X	14.00	0	0	0	0	0	0	0	9	0	14	0	0	0	100
Y	25.00	0	0	0	0	0	0	2	15	0	1	25	0	0	100
Z	0.00	0	0	0	0	0	0	0	0	0	7	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	27.26	82	0	0	223	0	168	246	163	0	55	104	6	102	2700

Table B.20: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 1 variable

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	16.00	16	0	0	0	7	0	0	0	12	0	0	0	0	2
B	12.00	0	12	0	0	0	10	26	0	0	0	7	[illegible]	0	0
C	9.00	0	0	9	42	0	1	0	4	0	0	0	23	0	0
D	57.00	0	0	3	57	0	1	0	2	0	0	0	21	0	0
E	83.00	1	0	0	0	83	0	0	0	0	0	0	0	0	0
F	28.00	0	4	0	3	0	28	22	0	0	0	0	1	0	0
G	33.00	0	10	0	1	0	17	33	0	0	0	0	0	0	0
H	1.00	0	1	4	43	0	5	3	1	0	0	0	12	0	0
I	18.00	12	0	1	0	[illegible]	0	0	0	18	0	0	0	0	3
J	77.00	0	0	0	0	[illegible]	0	1	0	0	77	13	0	0	0
K	55.00	0	4	0	0	0	0	1	0	0	7	55	0	0	0
L	30.00	0	0	7	41	0	2	1	1	0	0	0	30	0	0
M	0.00	0	0	2	27	0	14	5	5	0	0	0	10	0	0
N	2.00	15	0	1	0	2	0	0	0	8	0	0	0	0	2
O	20.00	7	0	1	0	3	0	0	0	20	0	0	1	0	8
P	0.00	0	0	0	15	0	17	12	5	0	0	0	2	0	0
Q	0.00	0	0	0	0	0	0	0	0	0	63	19	0	0	0
R	53.00	5	0	3	1	1	0	0	0	17	0	0	3	0	3
S	0.00	7	0	2	1	0	0	0	0	18	0	0	1	0	7
T	57.00	9	0	0	0	15	0	0	0	2	0	0	0	0	1
U	37.00	0	1	0	27	0	19	5	3	0	0	0	5	0	0
V	45.00	0	11	0	0	0	1	5	0	0	0	23	0	0	0
W	0.00	0	12	0	0	0	10	18	0	0	0	4	0	0	0
X	8.00	0	0	0	0	0	0	1	0	0	30	56	0	0	0
Y	24.00	0	16	0	0	0	12	28	1	0	0	2	0	0	0
Z	0.00	0	0	0	0	0	0	0	0	0	81	9	0	0	0
[illegible]	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	28.33	72	71	33	258	111	137	161	22	95	258	188	109	0	26

Table B.20 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	16.00	13	0	0	9	0	41	0	0	0	0	0	0	0	100
B	12.00	0	0	0	0	0	0	4	17	0	0	24	0	0	100
C	9.00	0	0	0	15	0	0	6	0	0	0	0	0	0	100
D	57.00	0	0	0	4	0	0	12	0	0	0	0	0	0	100
E	83.00	0	0	0	0	0	14	0	0	0	0	0	0	2	100
F	28.00	0	0	0	0	0	0	32	4	0	0	6	0	0	100
G	33.00	0	0	0	0	0	0	19	4	0	0	16	0	0	100
H	1.00	0	0	0	3	0	0	28	0	0	0	0	0	0	100
I	18.00	17	0	0	34	0	15	0	0	0	0	0	0	0	100
J	77.00	0	0	0	0	0	0	0	0	0	9	0	0	0	100
K	55[illegible]	0	0	0	0	0	0	0	25	0	4	4	0	0	100
L	30[illegible]	0	0	0	3	0	0	15	0	0	0	0	0	0	100
M	0.[illegible]	0	0	0	0	0	0	37	0	0	0	0	0	0	100
N	2.00	11	0	0	42	0	19	0	0	0	0	0	0	0	100
O	20.00	20	0	0	15	0	25	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	45	0	0	0	4	0	0	100
Q	0.00	0	0	0	0	0	0	0	2	0	15	1	0	0	100
R	53.00	12	0	0	53	0	2	0	0	0	0	0	0	0	100
S	0.00	15	0	0	36	0	13	0	0	0	0	0	0	0	100
T	57.00	9	0	0	7	0	57	0	0	0	0	0	0	0	100
U	37.00	0	0	0	0	0	0	37	0	0	0	3	0	0	100
V	45.00	0	0	0	0	0	0	0	45	0	0	15	0	0	100
W	0.00	0	0	0	0	0	0	6	22	0	0	28	0	0	100
X	8.00	0	0	0	0	0	0	0	4	0	8	1	0	0	100
Y	24.00	0	0	0	0	0	0	2	15	0	0	24	0	0	100
Z	0.00	0	0	0	0	0	0	0	0	0	10	0	0	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	28.33	97	0	0	221	0	186	243	138	0	46	126	0	102	2700

Table B.21: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	26.00	26	0	0	0	9	0	0	0	3	0	0	0	0	3
B	21.00	0	21	0	0	0	15	46	0	0	0	1	0	0	0
C	4.00	1	0	4	27	0	2	0	5	0	0	0	36	5	0
D	40.00	0	0	1	40	0	0	0	6	0	0	0	38	5	0
E	88.00	1	0	0	0	88	0	0	0	0	0	0	0	0	0
F	28.00	0	2	0	11	0	28	18	4	0	0	0	0	6	0
G	34.00	0	11	0	0	0	28	34	2	0	0	0	0	1	0
H	4.00	0	0	0	37	0	4	2	4	0	0	0	27	7	0
I	14.00	23	0	1	0	2	0	0	0	14	0	0	3	0	3
J	0.00	0	0	0	0	0	0	0	0	0	0	48	0	0	0
K	29.00	0	5	0	0	0	1	6	0	0	0	29	0	0	0
L	42.00	0	0	1	33	0	0	0	6	0	0	0	42	2	0
M	8.00	0	1	0	31	0	10	2	8	1	0	0	14	8	0
N	3.00	18	0	0	0	1	0	0	0	12	0	0	2	0	3
O	19.00	31	0	0	0	4	0	0	0	6	0	0	1	0	6
P	0.00	0	1	1	26	0	10	2	18	0	0	0	6	5	0
Q	0.00	0	0	0	0	0	0	0	0	0	0	46	0	0	0
R	[illegible]3.00	7	0	1	0	0	0	0	0	15	0	0	5	0	2
S	0.00	14	0	0	0	2	0	0	0	12	0	0	3	0	5
T	50.00	18	0	0	0	16	0	0	0	3	0	0	0	0	0
U	31.00	0	1	0	27	0	16	2	12	0	0	0	3	7	0
V	37.00	0	15	0	0	0	6	28	0	0	0	8	0	0	0
W	0.00	[illegible]	13	0	1	0	21	43	0	0	0	5	0	0	0
X	6.00	0	4	0	0	0	0	2	0	0	0	55	0	0	0
Y	0.00	0	11	0	0	0	21	45	0	0	0	1	0	0	0
Z	53.00	0	0	0	0	0	0	0	0	0	0	25	0	0	0
@	100.00	0	0	0	6	0	0	0	0	0	0	0	0	0	0
TOTAL	25.56	139	85	9	233	122	162	230	65	66	0	218	180	46	22

Table B.21 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	26.00	13	0	0	6	0	40	0	0	0	0	0	0	0	100
B	21.00	0	0	0	0	0	0	6	11	0	0	0	0	0	100
C	4.00	0	0	0	16	0	0	4	0	0	0	0	0	0	100
D	40.00	0	0	0	5	0	0	5	0	0	0	0	0	0	100
E	88.00	1	0	0	0	0	9	0	0	0	0	0	0	1	100
F	28.00	0	0	0	0	0	0	31	0	0	0	0	0	0	100
G	34.00	0	0	0	0	0	6	16	7	1	0	0	0	0	100
H	4.00	0	0	0	6	0	0	13	0	0	0	0	0	0	100
I	14.00	18	0	0	25	0	11	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	4	0	19	0	29	0	100
K	29.00	0	0	0	0	0	0	1	56	0	0	0	2	0	100
L	42.00	1	0	0	11	0	0	4	0	0	0	0	0	0	100
M	8.00	0	0	0	3	0	0	22	0	0	0	0	0	0	100
N	3.00	20	0	0	27	0	17	0	0	0	0	0	0	0	100
O	19.00	19	0	0	15	0	18	0	0	0	0	0	0	0	100
P	0.00	0	0	0	2	0	0	27	2	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	5	0	14	0	35	0	100
R	53.00	13	0	0	53	0	4	0	0	0	0	0	0	0	100
S	0.00	19	0	0	34	0	11	0	0	0	0	0	0	0	100
T	50.00	8	0	0	5	0	50	0	0	0	0	0	0	0	100
U	31.00	0	0	0	0	0	0	31	1	0	0	0	0	0	100
V	37.00	0	0	0	0	0	0	2	37	4	0	0	0	0	100
W	0.00	0	0	0	0	0	0	5	12	0	0	0	0	0	100
X	6.00	0	0	0	0	0	0	0	20	0	6	0	13	0	100
Y	0.00	0	0	0	0	0	0	13	7	2	0	0	0	0	100
Z	53.00	0	0	0	0	0	0	0	2	0	20	0	53	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	25.56	112	0	0	208	0	160	180	164	7	59	0	132	101	2700

Table B.22: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 3 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	30.00	30	0	0	0	8	0	0	0	7	0	0	0	0	3
B	6.00	0	6	0	0	0	19	47	0	0	0	2	0	2	0
C	5.00	0	0	5	32	0	0	0	0	0	0	0	41	2	0
D	48.00	0	0	0	48	0	1	0	5	0	0	0	35	3	0
E	88.00	2	0	0	0	88	0	0	0	0	0	0	0	0	0
F	25.00	0	3	0	3	0	25	19	2	0	0	0	2	7	0
G	36.00	0	4	0	0	0	33	36	1	0	0	0	1	3	0
H	10.00	0	0	1	40	0	2	3	10	0	0	0	19	6	0
I	14.00	26	0	0	0	0	0	0	0	14	0	0	0	0	4
J	0.00	0	0	0	0	0	1	0	0	0	0	42	0	0	0
K	30.00	0	11	0	0	0	1	6	0	0	0	30	0	0	0
L	45.00	0	0	0	34	0	2	0	2	0	0	0	45	2	0
M	10.00	0	0	0	35	0	8	3	12	0	0	0	13	10	0
N	0.00	21	0	0	0	2	0	0	0	11	0	0	1	0	0
O	23.00	21	0	0	0	3	0	0	0	13	0	0	1	0	5
P	0.00	0	0	0	17	0	17	6	12	0	0	0	5	14	0
Q	0.00	0	0	0	0	0	0	1	0	0	0	48	0	0	0
R	48.00	14	0	1	0	1	0	0	0	18	0	0	4	0	2
S	0.00	15	0	1	1	0	0	0	0	18	0	0	1	0	2
T	54.00	15	0	0	0	17	0	0	0	2	0	0	0	0	1
U	28.00	0	0	0	32	0	14	5	11	0	0	0	5	5	0
V	44.00	0	20	0	0	0	3	27	0	0	0	5	0	0	0
W	1.00	0	11	0	0	0	13	50	2	0	0	2	0	2	0
X	13.00	0	1	0	0	0	0	2	0	0	0	56	0	0	0
Y	0.00	0	10	0	1	0	26	46	0	0	0	1	0	0	0
Z	41.00	0	0	0	0	0	0	0	0	0	0	38	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	25.89	144	66	8	243	119	165	251	57	83	0	224	173	56	17

Table B.22 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	30.00	11	0	0	5	0	36	0	0	0	0	0	0	0	100
B	6.00	0	0	0	0	0	0	8	15	0	0	0	0	0	100
C	5.00	0	0	0	15	0	0	5	0	0	0	0	0	0	100
D	48.00	0	0	0	4	0	0	4	0	0	0	0	0	0	100
E	88.00	0	0	0	0	0	8	0	0	0	0	0	0	2	100
F	25.00	0	0	0	0	0	0	37	1	1	0	0	0	0	100
G	36.00	0	0	0	0	0	0	21	0	1	0	0	0	0	100
H	10.00	0	0	0	5	0	0	14	0	0	0	0	0	0	100
I	14.00	14	0	0	29	0	13	0	0	0	0	0	0	0	100
J	0.00	0	0	0	0	0	0	0	0	0	18	0	39	0	100
K	30.00	0	0	0	0	0	0	0	46	2	1	0	3	0	100
L	45.00	0	0	0	5	0	0	10	0	0	0	0	0	0	100
M	10.00	0	0	0	0	0	0	19	0	0	0	0	0	0	100
N	0.00	9	0	0	37	0	19	0	0	0	0	0	0	0	100
O	23.00	23	0	0	12	0	22	0	0	0	0	0	0	0	100
P	0.00	0	0	0	0	0	0	29	0	0	0	0	0	0	100
Q	0.00	0	0	0	0	0	0	0	5	0	19	0	27	0	100
R	48.00	11	0	0	48	0	1	0	0	0	0	0	0	0	100
S	0.00	23	0	0	26	0	13	0	0	0	0	0	0	0	100
T	54.00	6	0	0	5	0	54	0	0	0	0	0	0	0	100
U	28.00	0	0	0	0	0	0	28	0	0	0	0	0	0	100
V	44.00	0	0	0	0	0	0	0	44	1	0	0	0	0	100
W	1.00	0	0	0	0	0	0	6	13	1	0	0	0	0	100
X	13.00	0	0	0	0	0	0	0	19	0	13	0	9	0	100
Y	0.00	0	0	0	0	0	0	6	9	1	0	0	0	0	100
Z	41.00	0	0	0	0	0	0	0	0	0	21	0	41	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	25.89	97	0	0	191	0	166	187	153	7	72	0	119	102	2700

Table B.23: Classification matrix for training sample : quadratic discriminant, polyalphabetic, position 3, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	82.00	82	0	0	0	0	0	0	0	6	0	0	0	0	0
B	31.00	0	31	0	0	0	37	8	3	0	0	2	0	1	0
C	57.00	0	0	57	2	0	4	0	1	0	0	0	12	6	4
D	73.00	0	0	4	73	0	2	0	2	0	0	0	9	4	0
E	94.00	2	0	0	0	94	0	0	0	0	0	0	0	0	0
F	74.00	0	3	3	0	0	74	0	6	0	0	0	0	5	0
G	18.00	0	9	2	0	0	44	18	3	0	0	0	0	5	0
H	81.00	0	1	1	2	0	12	1	81	0	0	0	0	1	0
I	71.00	27	0	0	0	0	0	0	0	71	0	0	0	0	0
J	3.00	0	1	0	0	0	0	0	0	0	3	14	0	0	0
K	38.00	0	9	0	1	0	2	1	0	0	1	38	0	0	0
L	67 00	0	0	2	7	0	1	0	0	0	0	0	67	2	0
M	29.00	0	0	27	2	0	8	1	0	0	0	0	14	29	0
N	90.00	1	0	2	1	0	0	0	0	0	0	0	0	2	90
O	73.00	17	0	0	0	1	0	0	0	1	0	0	0	1	0
P	36.00	0	0	19	0	0	10	1	0	0	0	0	15	9	0
Q	52.00	0	0	0	0	0	0	0	0	0	1	3	0	0	0
R	70.00	1	0	2	0	0	0	0	0	0	0	0	3	0	9
S	92.00	0	0	0	0	0	0	0	0	1	0	0	2	0	0
T	86.00	7	0	1	0	0	0	0	0	0	0	0	0	0	1
U	75.00	0	1	0	0	0	1	1	1	0	0	1	11	3	0
V	52.00	0	7	0	0	0	6	6	0	0	0	2	0	0	0
W	24.00	0	26	0	0	0	35	7	0	0	0	7	0	0	0
X	33.00	0	4	0	0	0	0	0	0	0	1	21	0	0	0
Y	87.00	0	1	0	0	0	2	0	0	0	0	1	0	0	0
Z	80.00	0	0	0	0	0	0	0	0	0	0	5	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	61.78	137	93	120	88	95	238	44	97	79	6	94	133	68	104

Table B.23 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													Total
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	
A	82.00	10	0	0	0	0	2	0	0	0	0	0	0	0	100
B	31.00	0	5	0	0	0	0	1	3	9	0	0	0	0	100
C	57.00	0	5	0	6	0	1	2	0	0	0	0	0	0	100
D	73.00	0	2	0	1	1	0	0	1	0	0	1	0	0	100
E	94.00	2	0	0	0	1	1	0	0	0	0	0	0	0	100
F	74.00	0	8	0	0	0	0	0	0	0	0	1	0	0	100
G	18.00	0	10	0	0	0	0	2	3	1	0	3	0	0	100
H	81.00	0	0	0	0	0	0	0	0	1	0	0	0	0	100
I	71.00	1	0	0	0	0	1	0	0	0	0	0	0	0	100
J	3.00	0	0	11	0	0	0	0	0	0	12	0	59	0	100
K	38.00	0	0	2	0	0	0	0	29	7	3	2	5	0	100
L	67.00	0	3	0	10	1	0	7	0	0	0	0	0	0	100
M	29.00	0	15	0	0	0	0	4	0	0	0	0	0	0	100
N	90.00	0	0	0	3	0	1	0	0	0	0	0	0	0	100
O	73.00	73	0	0	1	2	4	0	0	0	0	0	0	0	100
P	36.00	0	36	0	1	0	0	6	0	1	0	2	0	0	100
Q	52.00	0	0	52	0	0	0	0	0	0	4	0	40	0	100
R	70.00	3	0	0	70	3	9	0	0	0	0	0	0	0	100
S	92.00	2	0	0	3	92	0	0	0	0	0	0	0	0	100
T	86.00	1	0	0	3	1	86	0	0	0	0	0	0	0	100
U	75.00	0	6	0	0	0	0	75	0	0	0	0	0	0	100
V	52.00	0	0	0	0	0	0	1	52	1	0	25	0	0	100
W	24.00	0	0	0	0	0	0	0	1	24	0	0	0	0	100
X	33.00	0	0	8	0	0	0	0	3	10	33	0	20	0	100
Y	87.00	0	2	0	0	0	0	0	7	0	0	87	0	0	100
Z	80.00	0	0	7	0	0	0	0	0	1	7	0	80	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	61.78	92	92	80	98	101	105	98	99	55	59	121	204	100	2700

Table B.24: Classification matrix for validating sample : quadratic discriminant, polyalphabetic, position 3, 20 variables

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP A	B	C	D	E	F	G	H	I	J	K	L	M	N
A	79.00	79	0	0	0	1	0	0	0	4	0	0	0	0	0
B	27.00	0	27	0	0	0	39	4	1	0	0	1	0	3	0
C	45.00	0	0	45	2	0	2	0	2	0	0	0	18	7	8
D	68.00	0	0	8	68	0	1	1	2	0	0	0	13	5	1
E	94.00	3	0	0	0	94	0	0	0	1	0	0	0	0	0
F	67.00	0	5	1	0	0	67	3	4	0	0	1	0	7	0
G	11.00	0	5	1	1	0	49	11	4	0	0	0	1	5	0
H	78.00	0	1	2	0	0	15	3	78	0	0	0	0	1	0
I	59.00	36	0	0	0	0	0	0	0	59	0	0	0	0	0
J	1.00	0	0	0	0	0	0	0	0	0	1	16	0	0	0
K	41.00	0	13	0	0	0	4	8	0	0	1	41	0	0	0
L	66.00	0	0	6	0	0	0	0	0	0	0	0	66	3	0
M	21.00	0	1	20	3	0	7	0	1	0	0	1	15	21	0
N	87.00	0	0	1	0	0	0	0	0	0	0	0	2	0	87
O	70.00	20	0	0	0	0	0	0	0	0	0	0	0	1	0
P	33.00	0	0	12	1	0	16	0	0	0	0	0	9	16	0
Q	40.00	0	0	0	0	0	0	0	0	0	1	11	0	0	0
R	72.00	5	0	1	0	0	0	0	0	0	0	0	3	1	8
S	88.00	2	0	0	1	0	0	0	0	1	0	0	1	0	0
T	81.00	9	0	1	1	0	0	0	0	0	0	0	0	0	1
U	70.00	0	0	0	0	0	0	0	1	0	0	0	18	2	0
V	55.00	0	2	0	0	0	2	2	0	0	1	4	0	0	0
W	16.00	0	28	0	0	0	45	2	3	0	0	1	0	0	0
X	30.00	0	5	0	0	0	0	0	0	0	3	21	0	0	0
Y	80.00	0	0	0	1	0	4	1	1	0	0	1	0	1	0
Z	67.00	0	0	0	0	0	0	0	0	0	1	12	0	0	0
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TOTAL	57.26	154	87	98	78	95	251	35	97	65	8	110	146	73	105

Table B.24 (continued)

GROUP	PERCENT CORRECT	NUMBER OF CASES CLASSIFIED INTO GROUP													
		O	P	Q	R	S	T	U	V	W	X	Y	Z	@	Total
A	79.00	14	0	0	0	0	2	0	0	0	0	0	0	0	100
B	27.00	0	9	0	0	0	0	0	0	15	1	0	0	0	100
C	45.00	0	6	0	9	0	0	1	0	0	0	0	0	0	100
D	68.00	0	1	0	0	0	0	0	0	0	0	0	0	0	100
E	94.00	1	0	0	0	1	0	0	0	0	0	0	0	0	100
F	67.00	0	11	0	0	0	0	0	0	0	0	1	0	0	100
G	11.00	0	20	0	0	0	0	2	0	0	0	1	0	0	100
H	78.00	0	0	0	0	0	0	0	0	0	0	0	0	0	100
I	59.00	5	0	0	0	0	0	0	0	0	0	0	0	0	100
J	1.00	0	0	15	0	0	6	0	0	3	10	0	55	0	100
K	41.00	0	0	8	0	0	0	0	18	3	1	1	2	0	100
L	66.00	0	10	0	6	0	0	9	0	0	0	0	0	0	100
M	21.00	0	23	0	0	0	0	8	0	0	0	0	0	0	100
N	87.00	0	0	0	6	0	4	0	0	0	0	0	0	0	100
O	70.00	70	0	0	1	5	3	0	0	0	0	0	0	0	100
P	33.00	0	33	0	0	0	0	12	0	1	0	0	0	0	100
Q	40.00	0	0	40	0	0	0	0	1	0	5	0	42	0	100
R	72.00	4	1	0	72	1	4	0	0	0	0	0	0	0	100
S	88.00	2	0	0	1	88	3	0	0	0	0	1	0	0	100
T	81.00	1	0	0	4	2	81	0	0	0	0	0	0	0	100
U	70.00	0	8	0	0	0	0	70	0	1	0	0	0	0	100
V	55.00	0	0	0	0	0	0	0	55	4	0	30	0	0	100
W	16.00	0	1	0	0	0	0	0	3	16	1	0	0	0	100
X	30.00	0	0	11	0	0	0	0	2	7	30	1	20	0	100
Y	80.00	0	3	0	0	0	0	0	8	0	0	80	0	0	100
Z	67.00	0	0	12	0	0	0	0	0	0	8	0	67	0	100
@	100.00	0	0	0	0	0	0	0	0	0	0	0	0	100	100
TOTAL	57.26	97	126	86	99	97	97	102	87	50	56	115	186	100	2700

APPENDIX C

Results from Multivariate Cryptanalysis

Figure C.1: Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 1 variable, Example 1

HETTETS LENT ELNTOT TMNOK GOH GOT TME NTTNCHE NO GOHT DNG NS
SHE NS COHTSE TNTGETEL NT DNONGETS YMNCM LESCTNWES TME CODMH
TET WNSEL NOGOTDNTNOO SGSTEDS MTOGTND NT TME HONAETSNTG OG A
NCTOTNN TMETE NS MOYEAET NO ETTOT NO TME TMNTL MNTNGTNMM OG
TMNT YTTNCHE CHTTEOTHG TMETE NTE NMMTOKNDNTEHG NCTNAE STHLEO
TS NO TME CODMHTET WNSLL NOGOTDNTNOO SGSTEDS MTOGTND NS OMMO
SEL TO TME STHLEOTS DEOTNOOEL NO TME NTTNCHE TME TOTNH OHDWE
T OG STHLEOTS YMO MNAE TEGNSTETEL GOT TME MTOGTND SNOCE NTS
NOCEMTNOO NO NS NMMTOKNDHTEHG OG TMOSE STHLEOTS MNAE CODMHET
EL TME MTOGTND TME CODMHTNOG TOOHS GOT DNONGEDEOT COHTSE LNS
CHSSEL NO TME NTTNCHE MNL NO EOTOHHDEOT OG STHLEOTS NS DEOTN
OOEL NT NS SODE OG ONOE COHTSES YMNCM CODMTNSE TME CWNS MTOG
TND SNOCETEHG KENOETTE DHKNO MTOGTND COOTLNONTOT CODMHTET WN
SEL NOGOTDNTNOO SGSTEDS HONAETSNTG OG ANCTOTNN ANCTOTNN WC H
ETTETS LENT ELNTOT N YNS MHENSEL TO OOTE SGLOEG LEAEHOMDEOT
COTMOTNTNOOS MTESS TEHENSE NOOOHOCNOG OHT NGTEEDEOT YNTM HON
SGS COTMOTNTNOO NO GOHT DNG NSSHE MOYEAET GOH MNL DE KHNCKHG
SMNKNOG DG MENL YMEO GOH STNTEL TMNT SGLOEGS DESSEOGET NOL
TGC GNTEYNG SOGTYNTE YOHHL WE NOCOTMOTNTEL NOTO HONSGSS EHEC
TTOONC DNSSNGNOG MTOLHCTS TNTMET TMNO DESSNGNOG MTOLHCTS NS
TMNS N OEY MTOLHCT HNOE YE NTE HONYNTE OG YMNHE MHDOTOHS SHC
M NO ETTOT NS N SOHTCE OG COOGHSNOO GOHTS SNOCETEHG WNTWNTN
DETHO DNTKETNOG DNONGET SGLOEG LEAEHOMDEOT COTM ANOCOHAET WC
HETTETS LENT ELNTOT N YNOT TO CHNTNGG OCT CODTEOS NOAOHAEDE
OT NO TME DOHOTNNO WEHH NSLO TTNNH NS LESCTNWEL WG N DNTCM C
ODMHTNOG CNONLN NTTNCHE NSLO ELGNOG CHOSET TO TENHNTG OCT CO
DTEO MTOANLEL CODDHONCNTNOOS MTOCESSOTS NOL OCT COTMOTNTNOOS
METSOONH CODMHTET LNANSNOO MTOANLEL METSOONH CODMHTETS GOT
DOHOTNNO WEHHS NSLO TTNNH OCT CODTEO NOL TME OCT MC LNANSNOO
LEAEHOMEL TYO NSLO TETDNONH NLNMTETS GOT TME TTNNH OOE NOTE
GTNTEL NOTO NO OCT MC TME SECOOL N STNOLNHOOE NLNMTET TMNS E
KHNMDEOT YNS HSEL TO DNKE TME GNTST CHSTODET NMMHNCNTNOO CNH
H OAET TME DOHOTNNO WEHH NSLO TTNNH OETYOTK LHTNOG TME OEYS
COOGETEOCE TMNT SNGONHEL TME WEGNOONOG OG TME TTNNH DNCMEHE
YOHGG MHWHNC TEHNTNOOS TEMTESEOTNTNAE COTMOTNTE CODDHONCNTNO
OS OCT CODTEO ST MNHH DNOO

Figure C.4. Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 3 variables, Example 1

HETTETS LENT ELNTOT TMNCX GOH YOT TME NTTNCHE NO GOHT DNG NS
SHE NS COHTSE TNTGETEL NT DNONGETS VMNCM LESCTNVES TME CODMH
TET VNSEL NOYOTDNTNCO SGSTEDS MTOGTND NT TME HONWETSNTG OY W
NCTOTNN TMETE NS MOVEWET NO ETTOT NO TME TMNTL MNTNGTNMM OY
TMNT NTTNCHE CHTTEOTHG TMETE NTE NMMTOJNDNTEHG NCTNWE STHLEO
TS NO TME CODMHTET VNSEL NOYOTDNTNOO SGSTEDS MTOGTND NS OMMO
SEL TO TME STHLEOTS DEOTNOOEL NO TME NTTNCHE TME TOTNH OHDVE
T OY STHLEOTS VMO MNWE TEGNSTETEL YOT TME MTOGTND SNOCE NTS
NOCEMTNOO NO NS NMMTOJNDNTEHG OY TMOSE STHLEOTS MNWE CODMHET
EL TME MTOGTND TME CODMHTNOG TOOHS YOT DNONGEDEOT COHTSE LNS
CHSSEL NO TME NTTNCHE MNL JO ECTOHHDEOT OY STHLEOTS NS DEOTN
OOEL NT NS SODE OY ONOE COHTSES VMNCM CODMTNSE TME CVNS MTOG
TND SNOCETEHG JENOETTE DHJNO MTOGTND COOTLNONTOT CODMHTET VN
SEL NOYOTDNTNOO SGSTEDS HONWETSNTG OY WNCTOTNN WNCTOTNN VC H
ETTETS LENT ELNTOT N VNS MHENSEL TO OOTE SGLOEG LEWEHOMDEOT
COTMOTNTNOOS MTESS TEHENSE NOOOHOCNOG OHT NGTEEDEOT VNTM HON
SGS COTMOTNTNOO NO GOHT DNG NSSHE MOVEWET GOH MNL DE JHNCXHG
SMNXNOG DG MENL VMEO GOH STNTEL TMNT SGLOEGS DESSEOGET NOL
TYC GNTEVNG SOYTVNTE VOHHL VE NOCOTMOTNTEL NOTO HONSGSS EHEC
TTOONC DNSSNGNOG MTOLHCTS TNTMET TMNO DESSNGNOG MTOLHCTS NS
TMNS N OEV MTOLHCT HNOE VE NTE HONVNTE OY VMNHE MHOOTOHS SHC
M NO ETTOT NS N SCHTCE OY COOYHSNOO GOHTS SNOCETEHG VNTVNTN
DETHO DNTXETNOG DNONGET SGLOEG LEWEHOMDEOT COTM WNOJOHWET VC
HETTETS LENT ELNTOT N VNOT TO CHNTNYG OCT CODTEOS NOWOHWEDE
OT NO TME DOHOTNNO VEHH NSLO TTNNH NS LESCTNVEL VG N DNTCM C
ODMHTNOG CNONLN NTTNCHE NSLO ELGNOG CHOSET TO TENHNTG OCT CO
DTEO MTOWNLEL CODDHONCNTNOOS MTOCESSOTS NOL OCT COTMOTNTNOOS
METSOONH CODMHTET LNWNSNOO MTOWNLEL METSOONH CCDMHTETS YOT
DOHOTNNO VEHHS NSLO TTNNH OCT CODTEO NOL TME OCT MC LNWNSNOO
LEWEHOMEL TVO NSLO TETDNONH NLNMTETS YOT TME TTNNH OOE NOTE
GTNTEL NOTO NO OCT MC TME SECOOL N STNOLNHOOE NLNMTET TMNS E
JHNMDEOT VNS HSEL TO DNXE TME YNTST CHSTODET NMMHNCHTNOO CNH
H OWET TME DOHOTNNO VEHH NSLO TTNNH OETVOTX LHTNOG TME OEVS
COOYETEOCE TMNT SNGONHEL TME VEGNOONOG OY TME TTNNH DNCMEHE
VOHYY MHVHNC TEHNTNOOS TEMTESEOTNTNWF COTMOTNTE CODDHONCNTNO
GS OCT CODTEO ST MNHH DNOO

Figure C.3: Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 12 variables, Example 1

```
METTETS DEOT EDITOT THONK YOP GOT THE OTTICME IN YOPT COY IS
SPE IS COPTSE TOTGETED OT CONOGETS WHICH DESCTIWES THE COCMP
TET WOSED INGOTCOTION SYSTECS MTOGTOC OT THE PNIAETSITY OG A
ICTOTIO THETE IS HOWEAET ON ETTCT IN THE THITD MOTOGTOMH OG
THOT OTTICME CPTTENTMY THETE OTE OMMTOQICOTEMY OCTIAE STPDEN
TS IN THE COCMPTET WOSED INGOTCOTION SYSTECS MTOGTOC OS OMMO
SED TO THE STPDENTS CENTIONED IN THE OTTICME THE TOTOM NPCWE
T OG STPDENTS WHO HOAE TEGISTETED GOT THE MTOGTOC SINCE ITS
INCEMTION IN IS OMMTOQICOTEMY OG THOSE STPDENTS HOAE COCMMET
ED THE MTOGTOC THE COCMPTING TOOMS GOT CONOGECENT COPTSE DIS
CPSSED IN THE OTTICME HOD ON ENTOMMCENT OG STPDENTS OS CENTI
ONED IT IS SOCE OG NINE COPTSES WHICH COCMTISE THE CWIS MTOG
TOC SINCETEMY ZEONETTE CPZIO MTOGTOC COOTDINOTOT COCMPTET WO
SED INGOTCOTION SYSTECS PNIAETSITY OG AICTOTIO AICTOTIO WC M
ETTETS DEOT EDITOT I WOS MMEOSED TO NOTE SYDNEY DEAEMOMCENT
COTMOTOTIONS MTESS TEMEOSE ONNOPNCING OPT OGTEECENT WITH PNI
SYS COTMOTOTION IN YOPT COY ISSPE HOWEAET YOP HOD CE QPICKMY
 SHOKING CY HEOD WHEN YOP STOTED THOT SYDNEYS CESSENGET OND
TGC GOTEWOY SOGTWOTE WOPMD WE INCOTMOTOTED INTO PNISYSS EMEC
TTONIC COSSOGING MTODPCTS TOTHET THON CESSOGING MTODPCTS IS
THIS O NEW MTODPCT MINE WE OTE PNOWOTE OG WHIME HPCOTOPS SPC
H ON ETTOT IS O SOPTCE OG CONGPSION YOPTS SINCETEMY WOTWOTO
CETMO COTKETING CONCTET SYDNEY DEAEMOMCENT COTM AONCOPAET WC
 METTETS DEOT EDITOT . WONT TO CMOTIGY NCT COCTENS INAOMAECE
NT IN THE COPNTOIN WEMM ISDN TTIOM OS DESCTIWED WY O COTCH C
OCMPTING CONODO OTTICME ISDN EDGING CMOSET TO TEOMITY NCT CO
CTEN MTOAIDED COCCPNICOTIONS MTOCESSOTS OND NCT COTMOTOTIONS
 METSONOM COCMPTET DIAISION MTOAIDED METSONOM COCMPTETS GOT
COPNTOIN WEMMS ISDN TTIOM NCT COCTEN OND THE NCT MC DIAISION
 DEAEMOMED TWO ISDN TETCINOM ODOMTETS GOT THE TTIOM ONE INTE
GTOTED INTO ON NCT MC THE SECOND O STONDOMONE ODOMTET THIS E
QPIMCENT WOS PS D TO COKE THE GITST CPSTOCET OMMMICOTION COM
M OAET THE COP. OIN WEMM ISDN TTIOM NETWOTK DPTING THE NEWS
CONGETENCE THOT SIGNOMED THE WEGINNING OG THE TTIOM CICHEME
WOMGG MPWMIC TEMOTIONS TEMTESENTOTIAE COTMOTOTE COCCPNICOTIO
NS NCT COCTEN ST MOPM CINN
```

Figure C.4: Cryptanalysis of Monoalphabetic Substitution by Using Quadratic Discriminant, 1 variable, Example 1

```
HETTEAN LENA ELNTAA TUNOX GAH YAA TUE NATNCHE NO GAHA DNG NN
NHE NN CAHANE TNAGETEL NT DNONGEAN VUNCU LENCANVEN TUE CADUH
TEA VNNEL NOYAADNTNAO NGNTEDN UAAGAND NT TUE HONWEANNTG AY W
NCTAANN TUEAE NN UAVEWEA NO EAAAA NO TUE TUNAL UNANGANUU AY
TUNT NATNCHE CHAAEOTHG TUEAE NAE NUUAAQNDNTEHG NCTNWE NTHLEO
TN NO TUE CADUHTEA VNNEL NOYAADNTNAO NGNTEDN UAAGAND NN AUUA
NEL TA TUE NTHLEOTN DEOTNAOEL NO TUE NATNCHE TUE TATNH OHDVE
A AY NTHLEOTN VUA UNWE AEGNNTEAEL YAA TUE UAAGAND NNOCE NTN
NOCEUTNAO NO NN NUUAAQNDNTEHG AY TUANE NTHLEOTN UNWE CADUHET
EL TUE UAAGAND TUE CADUHTNOG TAAHN YAA DNONGEDEOT CAHANE LNN
CHNNEL NO TUE NATNCHE UNL NO EOAAHHDEOT AY NTHLEOTN NN DEOTN
AOEL NT NN NADE AY ONOE CAHANEN VUNCU CADUANNE TUE CVNN UAAG
AND NNOCEAEHG QENOETTE DHQNA UAAGAND CAAALNONTAA CADUHTEA VN
NEL NOYAADNTNAO NGNTEDN HONWEANNTG AY WNCTAANN WNCTAANN VC H
ETTEAN LENA ELNTAA N VNN UHENNEL TA OATE NGLOEG LEWEHAUDEOT
CAAUAANTNAON UAENN AEHENNE NOOAHOCNOG AHA NGAEEDEOT VNTU HON
NGN CAAUAANTNAO NO GAHA DNG NNNHE UAVEWEA GAH UNL DE QHNCXHG
 NUNXNCG DG UENL VUEO GAH NTNTEL TUNT NGLOEGN DENNEOGEA NOL
AYC GNTEVNG NAYTVNAE VAHHL VE NOCAAUAANTEL NOTA HONNGNN EHEC
TAAONC DNNNNGNOG UAALHCTN ANTUEA TUNO DENNNGNOG UAALHCTN NN
TUNN N OEV UAALHCT HNOE VE NAE HONVNAE AY VUNHE UHDAAAHN NHC
U NO EAAAA NN N NAHACE AY CAOYHNNAO GAHAN NNOCEAEHG VNAVNAN
DEAHA DNAXETNOG DNONGEA NGLOEG LEWEHAUDEOT CAAU WNOCAHWEA VC
 HETTEAN LENA ELNTAA N VNOT TA CHNANYG OCA CADTEON NOWAHWEDE
OT NO TUE DAHOTNNO VEHH NNLO TANNH NN LENCANVEL VG N DNACU C
ADUHTNOG CNONLN NATNCHE NNLO ELGNOG CHANEA TA AENHNTG OCA CA
DTEO UAAWNLEL CADDHONCNTNAON UAACENNAAN NOL OCA CAAUAANTNAON
 UEANAONH CADUHTEA LNWNNNAO UAAWNLEL UEANAONH CADUHTEAN YAA
DAHOTNNO VEHHN NNLO TANNH OCA CADTEO NOL TUE OCA UC LNWNNNAO
 LEWEHAUEL TVA NNLO TEADNONH NLNUTEAN YAA TUE TANNH AOE NOTE
GANTEL NOTA NO OCA UC TUE NECAOL N NTNOLNHAOE NLNUTEA TUNN E
QHNUDEOT VNN HNEL TA DNXE TUE YNANT CHNTADEA NUUHNCNTNAO CNH
H AWEA TUE DAHOTNNO VEHH NNLO TANNH OETVAAX LHANOG TUE OEVN
CAOYEAEOCE TUNT NNGONHEL TUE VEGNOONOG AY TUE TANNH DNCUEHE
VAHYY UHVHNC AEHNTNAON AEUAENEOTNTNWE CAAUAANTE CADDHONCNTNA
ON OCA CADTEO NT UNHH DNOO
```

Figure C.5: Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 5 variables, Example 1

```
HETTEAN LEOA ELOTAA THOAK GAH GAA THE OATOCHE OA GAHA LOG ON
NHE ON CAHANE TOAGETEL OT LOAOGEAN WHOCH LENCAOVEN THE CALMH
TEA VONEL OAGAALOTOAA NGNTELN MAAGAOL OT THE HAOWEANOTG AG W
OCTAAOO THEAF ON HAWEWEA OA EAAAA OA THE THOAL MOAOGAOMH AG
THOT OATOCHE CHAAEATHG THEAE OAE OMMAAXOLOTEHG OCTOWE NTHLEA
TN OA THE CALMHTEA VONEL OAGAALOTOAA NGNTELN MAAGAOL ON AMMA
NEL TA THE NTHLEATN LEATOAAEL OA THE OATOCHE THE TATOH AHLVE
A AG NTHLEATN WHA HOWE AEGONTEAEL GAA THE MAAGAOL NOACE OTN
OACEMTOAA OA ON OMMAAXOLOTEHG AG THANE NTHLEATN HOWE CALMHET
EL THE MAAGAOL THE CALMHTOAG TAAHN GAA LOAOGELEAT CAHANE LON
CHNNEL OA THE OATOCHE HOL OA EAAAHHLEAT AG NTHLEATN ON LEATO
AAEL OT ON NALE AG AOAE CAHANEN WHOCH CALMAONE THE CVON MAAG
AOL NOACEAEHG JEOAETTE LHJOA MAAGAOL CAAALOAOTAA CALMHTEA VO
NEL OAGAALOTOAA NGNTELN HAOWEANOTG AG WOCTAAOO WOCTAAOO VC H
ETTEAN LEOA ELOTAA O WON MHEONEL TA AATE NGLAEG LEWEHAMLEAT
CAAMAAOTOAAN MAENN AEHEONE OAAAHACOAG AHA OGAEELEAT WOTH HAO
NGN CAAMAAOTOAA OA GAHA LOG ONNHE HAWEWEA GAH HOL LE XI CKHG
 NHOKOAG LG HEOL WHEA GAH NTOTEL THOT NGLAEGN LENNEAGEA OAL
AGC GOTEWOG NAGTWOAE WAHHL VE OACAAMAAOTEL OATA HAONGNN EHEC
TAAAOC LONNOGOAG MAALHCTN AOTHEA THOA LENNOGOAG MAALHCTN ON
THON O AEW MAALHCT HOAE WE OAE HAOWOAE AG WHOHE HHLAAAHN NHC
H OA EAAAA ON O NAHACE AG CAAGHNOAA GAHAN NOACEAEHG VOAVOAO
LEAHA LOAKETOAG LOAOGEA NGLAEG LEWEHAMLEAT CAAM WOACAHWEA VC
 HETTEAN LEOA ELOTAA O WOAT TA CHOAOGG ACA CALTEAN OAWAHWELE
AT OA THE LAHATOOA VEHH ONLA TAOOH ON LENCAOVEL VG O LOACH C
ALMHTOAG COAOLO OATOCHE ONLA ELGOAG CHANEA TA AEOHOTG ACA CA
LTEA MAAWOLEL CALLHAOCOTOAAN MAACENNAAN OAL ACA CAAMAAOTOAAN
 MEANAAOH CALMHTEA LOWONOAA MAAWOLEL MEANAAOH CALMHTEAN GAA
LAHATOOA VEHHN ONLA TAOOH ACA CALTEA OAL THE ACA MC LOWONOAA
 LEWEHAMEL TWA ONLA TEALOAOH OLOMTEAN GAA THE TAOOH AAE OATE
GAOTEL OATA OA ACA MC THE NECAAL O NTOALOHAAE OLOMTEA THON E
XHOMLEAT WON HNEL TA LOKE THE GOANT CHNTALEA OMMHOCOTOAA COH
H AWEA THE LAHATOOA VEHH ONLA TAOOH AETWAAK LHAOAG THE AEWN
CAAGEAEACE THOT NOGAOHEL THE VEGOAAOAG AG THE TAOOH LOCHEHE
WAHGG MHVHOC AEHOTOAAN AEMAENEATOTOWE CAAMAAOTE CALLHAOCOTOA
AN ACA CALTEA NT MOHH LOAA
```

Figure C.6: Cryptanalysis of Monoalphabetic Substitution Using Quadratic Discriminant, 12 variable, Example 1

```
LETTETS DEOT EDITOT THONK YOP GOT THE OTTICLE IN YOPT COY IS
SPE IS COPTSE TOTGETED OT CONOGETS WHICH DESCTIWES THE COCMP
TET WOSED INGOTCOTION SYSTECS MTOGTOC OT THE PNIVETSITY OG V
ICTOTIO THETE IS HOWEVET ON ETTOT IN THE THITD MOTOGTOMH OG
THOT OTTICLE CPTTENTLY THETE OTE OMMTOQICOTELY OCTIVE STPDEN
TS IN THE COCMPTET WOSED INGOTCOTION SYSTECS MTOGTOC OS OMMO
SED TO THE STPDENTS CENTIONED IN THE OTTICLE THE TOTOL NPCWE
T OG STPDENTS WHO HOVE TEGISTETED GOT THE MTOGTOC SINCE ITS
INCEMTION IN IS OMMTOQICOTELY OG THOSE STPDENTS HOVE COCMLET
ED THE MTOGTOC THE COCMPTING TOOLS GOT CONOGECENT COPTSE DIS
CPSSED IN THE OTTICLE HOD ON ENTOLLCENT OG STPDENTS OS CENTI
ONED IT IS SOCE OG NINE COPTSES WHICH COCMTISE THE CWIS MTOG
TOC SINCETELY QEONETTE CPQIO MTOGTOC COOTDINOTOT COCMPTET WO
SED INGOTCOTION SYSTECS PNIVETSITY OG VICTOTIO VICTOTIO WC L
ETTETS DEOT EDITOT I WOS MLEOSED TO NOTE SYDNEY DEVELOMCENT
COTMOTOTIONS MTESS TELEOSE ONNOPNCING OPT OGTEECENT WITH PNI
SYS COTMOTOTION IN YOPT COY ISSPE HOWEVET YOP HOD CE QPICKLY
 SHOKING CY HEOD WHEN YOP STOTED THOT SYDNEYS CESSENGET OND
TGC GOTEWOY SOGTWOTE WOPLD WE INCOTMOTOTED INTO PNISYSS ELEC
TTONIC COSSOGING MTODPCTS TOTHET THON CESSOGING MTODPCTS IS
THIS O NEW MTODPCT LINE WE OTE PNOWOTE OG WHILE HPCOTOPS SPC
H ON ETTOT IS O SOPTCE OG CONGPSION YOPTS SINCETELY WOTWOTO
CETLO COTKETING CONOGET SYDNEY DEVELOMCENT COTM VONCOPVET WC
 LETTETS DEOT EDITOT I WONT TO CLOTIGY NCT COCTEN INVOLVECE
NT IN THE COPNTOIN WELL ISDN TTIOL OS DESCTIWED WY O COTCH C
OCMPTING CONODO OTTICLE ISDN EDGING CLOSET TO TEOLITY NCT CO
CTEN MTOVIDED COCCPNICOTIONS MTOCESSOTS OND NCT COTMOTOTIONS
 METSONOL COCMPTET DIVISION MTOVIDED METSONOL COCMPTETS GOT
COPNTCIN WELLS ISDN TTIOL NCT COCTEN OND THE NCT MC DIVISION
 DEVELOMED TWO ISDN TETCINOL ODOMTETS GOT THE TTIOL ONE INTE
GTOTED INTO ON NCT MC THE SECOND O STONDOLONE ODOMTET THIS E
QPIMCENT WOS PSED TO COKE THE GITST CPSTOCET OMMLICOTION COL
L OVET THE COPNTOIN WELL ISDN TTIOL NETWOTK DPTING THE NEWS
CONGETENCE THOT SIGNOLED THE WEGINNING OG THE TTIOL CICHELE
WOLGG MPWLIC TELOTIONS TEMTESENTOTIVE COTMOTOTE COCCPNICOTIO
NS NCT COCTEN ST MOPL CINN
```

Figure C.7: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 1 variable, Example 1

```
123123123123123123123123123123123123123123123123123123123123

LETEETO CEOT ECROTI TLOTW GIH UTT OGE TIOTCAE OT UTLT COG OR
OHE OR CTLTRE IRTWEOEL TE HRITGETO GLOCL CESCIOGES ELE CTCCU
EET GTOEC OTUTTCOTRTT SGOOECS UTTGTTC TE TLE LITGETOOTG TU F
RCTITTR TLETE TO GIYEGET OT ETITT OT OGE TLOTL MRTTGTTUL IY
ELTE TIOTCAE CUITEOOHG TLETE TIE RCMITVRMTEEHG TCOTGE OOULET
ES RI ELE CTCCUEET GTOEC OTUTTCOTRTT SGOOECS UTTGTTC TO TUCT
OEC OT OGE REHCEITO HEI1RTTEH RI ELE OTEUCLE ELE OTEOH IUCGE
I TU REHCEITO GLT LOFE TEUTOOEIEC YTI TLE UTTGTTC RRICE TES
RICECTRTT OT OR OMUTTVOHROELU IY ELTOE COULETES LOFE CIMMLET
EH ELE CTIUTRM ELE CTCCUEOTG TITHO GIT COTRUECETE CIHTOE LOR
CHROEC OT OGE TIOTCAE LTL TO EOTTLAHEIT TG STLHEOOR OR MEOOT
IIEL TE TO RIME TG ITOE CTUISEO GLOCL CIMMIORE TLE CGTO MITW
IOH STOCEIEHG VEOTEOTE HLVTI MITWIOH CTITCRITETT CTCCUEET GT
OEC OTUTTCOTRTT SGOOECS LITGETOOTG TU FRCTITTR FRCTITTR GC H
EOTETR HERT EHTETT O GOR CHEOREH ET OTTE RGHTEU LEFEATUMEOO
CTTUTTROTIIR CTESR TELETOE RITIHTCOTG TLT RUTEEHEIT YTEL LIT
OUR CTICTIOTRTT OT UTLT COG OROHE LTGEFET GTU LTL HE VLOCWAG
 SGRVTOU CU LETL GLET UTL REOTEH ELTE RGHTEUR MEOSEOUEI TOH
IYC UTEEGRU OTGEYTIE GTULH VE RICITMITTEEC OTET LITOURO ELEC
ETTOOC MTOSTGOTG MITCLCTO TROGET ELTO HESRRUTOU UTTLHCES RS
ELTO T IEC MITCLCT ATOE GE RTE HTRYTIE IY GLTLE LHHITTLS OHC
L TO EITTI TO T STLTCE TU CIIGLSTII GTUIS OOTCETEAG GTIGTIO
CETLT COTWETRIW MTOOWET OUCOEG HEGEHICHEIT CTIC GOTCTUGET GC
 AEEOEIS LETI ELOTIT R GRIT OT CHRTTUU OCT CTCOEOS RIFIAFEME
OO RI ELE MTLITROT GELA RSCG TIOTL TO CESCIOGEH VU R HRTCL C
IMMLOTOU COTRHT OTEOCLE RSCO ELUTOU CATOET OT TERATEU OCT CT
COEO MITFRHEL CIMHLITCOTRTTO MITCESRITR OTL TCT CTTUTTROTIIR
 CEISTOOH CTCCUEET HTGORRTT CTIVTLEC CEISTOOH CTCCUEETO GIT
CTUOOTRI VEHLS RSCO TIOTL TCT CTHEET OTL TLE OCT CC HTGORRTT
 HEGEHICEL TGT RSCO TETHRITL TLOMEETO GIT ELE OTROH TTE TOOE
GTTEEC OTET RI OCT CC OGE RECTOH R REOTLOHIIE OCRCTET ELTO E
VHTUMEOO GOR HREH ET COVE TLE UOTOO CHRETHET RCMLOCROTII COH
L TGET OGE HIHTEOTO GEAH ORLI ETTRA OETGTTW CLTTOU ELE IEGS
CTTUETEICE TLOT STGITLEC OGE GEUTOITOU IY ELE OTROH MTCLELE
GTHUY UHGLOC TELOTRTTO TECTESEOOTEOFE CITMITTEE CTHCHTRCTEOT
OS OCT CTCOEO RE MRHH MTOI
```

Figure C.8: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 3 variables, Example 1

```
123123123123123123123123123123123123123123123123123123123123

IGULGXGAXEXXIGXLEMXAUKXTXAZCYTF XIEXEAXXEG TTIOTIZMRHTRXZIOT
GYTIOTIZMRHTEAUYHVEETLAXLATYXXZGXGAWKOHKAXEHHXOHEHTLLTIZMRSG
LGXIKXGGXIOTF XRXUL TIHZGETRHTGHMZHXRAXLAUKGTRXGOGXGOUCAMFAE
LZUCHGYAUKGXEAGGAXCATOGXIXTIGXX XIOTIEXEAUKOXLAGYHXZHXGLTCCT
LLXLAXXEG TTIZGXHTKECCAUKGXEAXXGTYSGX VLUXLGCCAX EGOGTGEGLGT
LHTLXTLLTIZMRSGLGXIKXGGXIOTF XRXUL TIHZGETRHTGHMZHXRAXGAMGSM
GGXIEMIEXEATLYXEXUGATEXUL TEOTLXTLLTIXXLOHIGTLLTIEMLXCIXGRKT
XAMFATLYXEXUGAWK TKXEEAXEUGGETXGXICMXAUKGTGHMZHXRATLXHEAGLHT
LXHESUL TIOTIOTIXGGHMLOTYETIZTCCTLLMGGTGEGLGTLHTKXEEAHCUGIGU
EOTLLTISXCUXYUTLLTIZMRSGLOTZAUC CGARCHTRXTYUTRGTLAHCYXGGTLOT
 YT    OTIEXEAXXEG TTILXLAXKATKHMITTEXUI RIHUROTKETIXTIUTKEG
CXTL   AGGATCUTI RIXGKGT  GXHTGAWKOHKAHCUGXOTEAUKGT KGGAGX V
XXTIHG ZTXGCCAAEXTEEUEATRXGCAGX VXXTIZMCHXLXXL XIZMRSGLGXIKX
GGXIOTF XRXUL TIHZGETRHTRXGOGXGOUCAMFAELZUCHGYAELZUCHGYAH AC
EEUEHTIOTYHTEOGL XIOTEXTISCEXTEOTL TK UEATCOTEZTLGEETMGUTKET
  XG XYEGCXTISXEHTIHTIGXGGTYXTCYT OTZAMRHTYUXEGTEXUIAGLLTRXG
GZTIZMXSMXXUL TIOTIZMRHTRXZIOTGYTILMEGEEHTC GILXLATEATROHXTZ
IHXYIGKUTRZTKGXLAWKGTIZMRATLXUEOTLLXLATCOTEZTIUTGHTKUTXAXKOT
XCHIUXLGWYZTG RLAXXGTE GIOTZGTLXHCHGCHXLGXIOTL TRXGGZTGATIGH
LHMKOHIUXGHXZOTZAGX XRZUGAXYEXEHTLLXKATEHTYUGKUTGHMLYHLHTLHT
LLGGAXIXTEAGX XRZUITGKGTEGTYHTIYTYAXXGTCCTELGIGTKYTCHMRHTGYH
KAXKATXHMXAGGAXIHMRHHEAMFAHCXRRHGCXTC GXHTGOT GXETZIKXXKXXXT
RGXI TRXXXGULXVIUXKXVEHTGZXKGZIOTOGCCSTEXUIZMXSTOXT  GOGXIKH
ITTLETXHTLGXXATLOUCHTLAWYXUIEMIZCYHGFZTKZXIZMRETKHTLXECTEEUT
KETLXTLLTIUMRXUYOTIKTITTLHXKAUXOXIAXGAXEHHXOHEOTZZTYATYHHKAH
CUGREGKUT XTYOXIXXLOHIGTLHXKATLUGKUT TMGGXIEMIHTYTGLZTKZXIZM
RETKAGX ELOTLAHCUTRXG XUL TGAGX HEHTCHTIXTLAT  T  XG XYEGCXT
ISTXHMKXCIZMRSGLGXIOGOOTL TISXCCGLGXISTXHMKXCIZMRSGLGXGARCHT
R GKEXLXTZGCIHTLHXKAUXOXIAT HT  TLGTIXTLAUKGTKZXISHIOGOOTL T
IOTOGCCSTLAUE TLHXKAUEHTLXXIAXLXGLGXGARCHTLLTIEXLXCI TEAGKET
ZHXLGXIOTL TYXTKZXISHIEXEATEZMKOTYATLXTLXCCXTIXXYSUEHTLLGGAT
GYGGUTKETEXTIYTEOTL TRXCEAUKGTFOXGET YTL TEHTYSGIOHYEGCXT XC
IAMOGXIEXEATCYTLXGKAHETCIOTLXTLHGYTTKGUS YXAXRHGKUTLLTIXTEHT
  TFGXEXHEAUKXUIHGZXXIGXIEXEAHEUGKXGKUT   LLTIEXLXCIUG LTIGT
E CFCTGYHIOHIHTIXUL TGAXESXEHTKEXLOEEAHCHGCHXLGT  TRYTLZXLOM
KHTKZXIZMRETKATLAGYYCIUGKXT
```

Figure C.9: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 20 variables, Example 1

```
123123123123123123123123123123123123123123123123123123123123

LETTENS DERN EDCINR TDRNK YSW FON IYE TRIICPE SN DNLR CRY SR
SWE SR CNLRRE TOPVEIEL TT UORTGENS WDSHD DERHRSAER THE CNCLM
TEN FTSED SNFONCRTCON RYSIECR GRNGRTC TT TDE LRIYENSSTW NF Y
CCTSRIO TDENE IS YSUEYEN RN ENRON SN IYE TDSNL FORTGRTGH SY
THTT TRIICPE CMRRENIHW TDENE TRE OLFROXCFTTEHW TCIIYE SIMLEN
TR CR THE CNCLMTEN FTSED SNFONCRTCON RYSIECR GRNGRTC TS NGLN
SED IN IYE RTWDERTS UERTCONED CR THE RNTSHLE THE INTRH RMCFE
R NF RTWDERTS WDO DRYE NEHISIERED YNR TDE GRNGRTC RCRHE ITR
CRHELTCON SN SR RFGRNKSUOIELD SY THNSE SIMLENTR DRYE HSFFLET
ED THE LNSHNOF THE CNCLMTSNG TSOHS GSR CRNOHECENT HSWNSE LSR
CWRSED SN IYE TRIICPE HTL TN ENRNLPUERT OG RTLDENIR RR FENII
SREL IT IS RSFE OG RINE COMRRES WDSHD HSFFRSRE TDE CFIS FROV
RRU RINCEREHW JERNEITE ULKIS FROVRRU CNSRDCRTTON CNCLMTEN FT
SED SNFONCRTCON RYSIECR LRIYENSSTW NF YCCTSRIO YCCTSRIO AC H
EITERR DEOR EDITON S PRR LHERRED TO NOTE RWDNED LEYEPNGFENI
CONGONOIISRR LNERR RELETSE ORNSWNCSNG NLR OHNEEUERT UITH LRI
SDR CNRINRRTCON SN DNLR CRY SRSWE HNPEYER WOM HTL UE KLSHKPY
 RYOKINH CD DETL WDEN DNL RTRTED THTT RWDNEDR FESRENHER TND
RYH HTTEWOD SOGTUTRE POMLD XE CRHSRFSRTTED SNTO LRISDRS ELEH
TRNNSH FTSRTGSNG FRODLCTS NOIYER THTN UERROHINH GRNLWHTR CR
THIS T REP FRODLCT PINE PE ORE WNOUTRE SY PHILE DWUSRNLR SWH
D TN E RNR IS T RNLRHE NF HSRGLRISR WOMRR SSNCENEPY FTRFTRR
CENLO CRNKETCRV FTNRVER SDDNEY DEYEHSLUERT CNRL YRNCOMYEN FH
 PETIERR LETR ELSTSR C WORT IN CHORIFD NCN CNCIENR CRYSPYEFE
NI CR THE FNLRTOSN FELP CRDN TRSTL TS DERHRSAED XD O UORHD H
SFFLIINH CRNODT RNTSHLE CRDN ELHINH CPNSEN IN REOPITD NCN CN
CIEN FROYCDEL HSFULRICRTCONS FROHERRSRR RNL NCR CONGONOIISRR
 LERRNNRH CNCLMTEN DIYSRCON LNSJILED LERRNNRH CNCLMTENS GSR
COMNITCR XEHLR CRDN TRSTL NCR COUTEN RNL TDE NCN LH DIYSRCON
 DEYEHSLEL TPO CRDN TERUCRTL TLRFTENS GSR THE INCRH ONE INIE
GRTTED SNTO OR NCN LH IYE RECNND O RTRNLRHSRE RDOLTER THIS E
KWIGFENI PRR WRED TO CRXE TDE FSNSI CWRTOUER OLFLSHOIISR CRH
L NYEN IYE USWNTRIN AEPH SRLR TRIOP NETPONK DLRINH THE REPR
CONFENERHE TDRT RIGRTLED IYE AEHINRINH SY THE INCRH FICHELE
POHFY GWALSH PELRTCONS NELNERENITTSYE HSRFSRTTE COUCWNCCTTSN
NR NCN CNCIEN RT FOWH FINR
```

Figure C.10: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 1 variable, Example 1

```
123123123123123123123123123123123123123123123123123123123123123

LETEETA LEOT ELROTS TLOTV YSH GTT OGE TSOTLUE OT UTLT COY OR
AHE OR CTLTRE TRTVEOEL TE DRITGETA VLORL LERRSOVER EDE CTCLU
EET YTAEL OTGTTCOTRTT RYAOECR GTTGTTC TE TLE LITGETAOTY TG G
RCTSTTR TLETE TA GSKEGET OT ETSTT OT OGE TLOTL URTTGTTGD SV
EDTE TSOTLUE CUSTEAODY TLETE TSE RLUSTJRUTEEDY TLOTGE AOULET
ER RI EDE CTCLUEET YTAEL OTGTTCOTRTT RYAOECR GTTGTTC TA TGLT
AEL OT OGE REHLEITA DEITRTTED RI EDE OTEORLE EDE OTEOD IUCYE
S TG REHLEITA VLT LOGE TEUTAOESEL VTS TLE GTTGTTC RRIRE TER
RIRELTRTT OT OR OUGTTXODROELU SV EDTAE AOULETER LOGE RSUULET
ED EDE LTSUTRU EDE CTCLUEOTG TSTDA BST COTRUECETE RSHTAE LOR
LHRAEL OT OGE TSOTLUE DTL TA EATTLUDEIT TB RTLDEAOR OR UEAOT
SIEL TE TA RSUE TB ITAE LTUSREA VLORL RSUUSORE TLE 'YTA USTV
SOD RTACESEDY JEOTEOTE DLXTS USTVSOD CTSTLRITETT CTCLUEET YT
AEL OTGTTCOTRTT RYAOECR LITGETAOTY TG GRCTSTTR GRCTSTTR VL D
ECTETR DERT EDTETT O GOR LDEORED ET ATTE RYDTEU LEGEUTGUEAO
LTTGTTROTSIR LTERR TELETAE RITSHTLOTG TLT RUTEEDEIT KTED LIT
AUR CTSLTSOTRTT OT UTLT COY ORAHE DTGEGET YTU DTL DE JLORVUY
 RGRXTAU CU LETL VLET UTL REOTED EDTE RYDTEUR UEAREAUES TAD
SVR UTEFVRU ATBEKTSE GTULD KE RIRSTUSTTEEL OTET LITAURA ELER
ETTAOR UTARTGOTG USTLLCTA TROGET EDTA DERRRUTAU GTTLHRER RR
EDTA T IEG USTLLCT UTAE GE RTE HTRKTSE SV GDTLE LHDSTTLR AHR
L TA ESTTS TA T RTLTRE TG RSIBLRTSI YTUSR AOTLETEUY YTSYTSO
CETLT COTVETRIV UTAOVET AULAEY DEGEDSLDEIT CTSL GOTLTUGET YR
 UEEOESR LETS ELOTST R VRIT OT CDRTTGU ACT CTCOEAR RIGSUGEUE
AO RI EDE UTLITROT YELU RRLA TSOTL TA LERRSOVED KU R DRTRL R
SUULOTAU LOTRDT OTEORLE RRLA ELUTAU LUTACT OT TERUTEU ACT CT
COEA USTGRDEL RSUDLITLOTRTTA USTRERRSTR OTL TLT LTTGTTROTSIR
 LESRTAOD CTCLUEET D..."RRTT LTSXTLEL LESRTAOD CTCLUEETA BST
CTUAOTRI KEDLR RRLA TSOTL TLT LTDEET OTL TLE ACT LR DTGORRTT
 DEGEDSLEL TGT RRLA TETDRITL TLOUEETA BST EDE OTROD TTE TAOE
GTTEEL OTET RI ACT LR OGE RECTAD R REOTLODSIE OLRLTET EDTA E
XHTGUEAO GOR HRED ET COJE TLE GOTAO LHRETDET RLULORRCTSI LOD
L TGET OGE DSHTEOTA VEUD ORLI ETTRU AETGTTV LLTTAU EDE IEGR
LTTGETEIRE TLOT RTGITLEL OGE VEUTAITAU SV EDE OTROD UTLDELE
GTDGV GHVLOR TELOTRTTA TELTEREAOTEOGE RSTUSTTEE LTDCHTRCTEOT
AR ACT CTCOEA RE URHD UTAI
```

Figure C.11: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 3 variables, Example 1

```
123123123123123123123123123123123123123123123123123123123123123

LEAEETA LEO. ELRAAN ALOAV GND FTT AGE TNATLUE OA UALT COG OR
ADE OR CALTRE ARTVEAEL TE DROTFETA BLORL LEIRNOBEI ELE CACLU
EET GTAEL OAFTTCOARTA IGAAECI FTAFTTC TE ALE LOTFETAOAG AF F
RCANTTR ALETE TA GNVEFET OA ETNTT OA AGE ALOTL MRTTFTTFL NW
ELTE TNATLUE CUNTEAADG ALETE TNE RLMNTKRHTEEDG TLATFE AAULEA
EI RO ELE CACLUEET GTAEL OAFTTCOARTA IGAAECI FTAFTTC TA AFLA
AEL AA AGE REDLEOAA DEOARTAED RO ELE OTEORLE ELE AAEOD OUCGE
N AF REDLEOAA BLT LOFE TEUTAAENEL WAN ALE FTAFTTC RRORE TEI
RORELARTA OA OR OMFTAKODRAELU NW ELAAE AAULEAEI LOFE RNHMLEA
ED ELE LTNUTRH ELE CACLUEOAF ANTDA GNT COARUECEAE RNDTAE LOR
LDRAEL OA AGE TNATLUE LTL TA EATALUDEOA TG IALDEAAR OR HEAAT
NOEL TE TA RNHE TG OTAE LTUNIEA BLORL RNHMNORE ALE LGTA MNTV
NOD ITACENEDG KEOAEAAE DLKTN MNTVNOD CANTLROTETT CACLUEET GT
AEL OAFTTCOARTA IGAAECI LOTFETAOAG AF FRCANTTR FRCANTTR BL D
EAAETR DERT EDTETT O FOR LDEORED ET ATAE RGDAEU LEFEUAFHEAA
LTTFTTRATNOR LTEIR TELETAE ROANDALOAF ALT RUTEEDEOA VTEL LOT
AUR CANLANOARTA OA UALT COG ORADE LAFEFET GTU LTL DE KLORVUG
 IGRKTAU CU LETL BLEA UAL REOAED ELTE RGDAEUR HEAIEAUEN TAD
NWR UTEEBRU ATGEVTNE FTULD VE RORNTMNTTEEL UAET LOTAURA ELER
ETAAOR HTAITFOAF MNTLLCAA TRAGET ELTA DEIRRUTAU FTALDRFI RI
ELTA T OEF MNTLLCA UTAE FE RTE DARVTNE NW FLTLE LDDNTALI ADR
L TA ENTAN TA T IALTRE AF RNOGLITNO GTUNI AOALETEUG GTNGTNO
CETLT COTVEAROV HTAOVET AULAEG DEFEDNLDEOA CANL FOALTUFET GR
 UEZAENI LETN ELOANT R BROA AA CDRTTFU ACT CACAEAI ROFNUFEHE
AA RO ELE HALOAKOA GELU RILA ANOTL TA LEIRNOBED VU R DRTRL R
NHMLATAU LOARDT OTEORLE RILA ELUTAU LUAAET AA TERUTEU ACT CA
CAEA MNTFRDEL RNHDLOTLOARTAA MNTREIRNTR OAL ALT LTTFTTRATNOR
 LENIAAOD CACLUEET DTFORRTA LTNKTLEL LENIAAOD CACLUEETA GNT
CTUAATRO VEDLI RILA ANOTL ALT LTDEEA OAL ALE ACT LR DTFORRTA
 DEFEDNLEL AFT RILA AETDROTL TLOMEETA GNT ELE ATROD TAE TAAE
FTTEEL OAET RO ACT LR AGE RECAAD R REOALODNOE OLRLAET ELTA E
KDTFHEAA FCA DRED ET COKE ALE FOTAA LDRETDET RLMLORRATNO LOD
L AFET AGE DNDAEOTA BEUD ORLO ETTRU AEAFTTV LLTTAU ELE OEFI
LTAFETEORE ALOA ITFOTLEL AGE BEUTAOTAU NW ELE ATROD HTLLELE
FTDFW FDBLOR TELOARTAA TELTEIEAATEOFE RNTMNTTEE LTDCDARCTEOA
AI ACT CACAEA RE MRDD HTAO
```

Figure C.12: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 20 variables, Example 1

```
123123123123123123123123123123123123123123123123123123123123123

LHNWHSRCMMBSKHMMAWRCNGBAWCYRB FRSKAGMCARAPCUAKRAKYWLR BBYKRC
RBAKRCKHWLRCMCNRRVMAAUCAWCURRAPHSRCBGRWGCMMHWRRYMH WHAKHWBLP
WHSKGARHMKRAFRSBBNMRAKHYRAABH FRWPRABCAWCNGH LRPYHSRRNGCWFCY
MHNRRPRCNGHSMCPRCGRBAYHSKBAKHSRRSKRAKAGMCNGRSUCDRRAPRAFH RY
WHAWCARAPCUAKHPRRABAWGCNGHSMCARH RLDRRQMHAWHWGCACAPYH RAPUHA
WH MR WHAKHWBLPWHSKGARHMKRAFRSBBNMRAKHYRAABH FRWPRABCARCWFLW
RHMKAWKAGMCCWBMMRNRCUMRNMRAMD MR WHAKBSWRWLH WHAKAWWBWKRPBGA
RCWFCCWBMMRNRCBGR GBYMCSMHPRAARHMKYWRCNGH FRWPRABCCMRWMCPWH
MRWMLNMRAKRAKRCKBDFRWKRURAALY RY WHWRH RAPUHAWH GBYMCWRHDLHN
MD WHAKLSRHSRH WHAKHWBLPWRAPCNRRWRCVRR BBARHABHAWCWRBSRH URC
CBCRHMKRAKAGMCARAPCUAKHAUCABCABRWLUUMRNKKVKHNLDABACKBCKHABAP
RRAUCPWCPRCCRHAKRVKRPBH CRPRHARCBGRWGCWRHDRRCMCNGH CGPRCDRRV
RBUKHPBHARHWGCZMRAMANMCULKPRCDRRVRBUKHWRRMMRAWRSKHWBLPWHSKGA
RHMKRAFRSBBNMRAKHYRAABH LRPYHSRRNGCWFCYMHNRRPRCYMHNRRPRCYCCW
MANMRCKDARR MDPWRSKR PBCKLWMBCMD WR BRNMCCGDAMY UHYMUWFHABA
CRSFRSRAPRRCKLSMHCKRALHARH RRARBACRAPCWLR RHSMHUMRNKBPWH LRP
RYCKHWRLWRBNMRAKRAKYWLR BBYKRCRBAKHWPHYMR GRPKHAUCUMCKLRWWUY
KHGRKPBH BY GHAUCBGHAKYWLCCWBNMD WHAWCCGDAMYCKHARHABHARCABD
RYWKHAWHBRY RRVWBARH FRPLD XH MRWRRDRRAWHMKRAWR LRPRYCRCALHW
WRWBRWKHARHAPRAPCDRRMLHNRCSRAGMR WHABCUMHCRHPBH FRWUBWWH MH
WHPRCAKRAPCDRRMLHNKUPBH PH RRAKBARBARH RY PHPLH GBURRWLH RBW
GCABCARRWRCPRCAKHWLRWMCWFCWRRVLHPRR GRPRH RRACHSMUYKGARGARB
BHSLR BBSWHNMRVKHABBVMR RYMBHYKDAYHWRLUMRNKHWRL YBACRPYHSKGW
KUAWAARH UHARCAURNRR MCBRRNKAWKHWRRPFY BHSKHWBAABH MRYRUYMHA
EA MR WHAKHWLRNRRAKGALU MHMBCNRRALCARCMMHWRRYMD XY RCURRWGCW
RHDLAPBH CBARDAKBSWRWLH MHMBCAUHPBH CUWRHSKAWKRARUPWY BHSKHW
BAABCDRRYMDAUCWRHULRPCBNMRARCDRRWMHCRRCKBAUCACR CRSFRSRAPRRC
KLARHWBBWKHWBLPWHSKDPYRCMRAKLSRZPUHMKLARHWBBWKHWBLPWHSRCVRR
BRPBAAMR XHWLH MHMBCNRRALCACR CRUWHAKBAUCNGH BHSKLWKDPYRCMRA
KDAYHWRLAUCNPR MHMBCNMRUMRALCAUBDWHSRCVRR WHAKASMBWKRAMCPBAA
PRAWHMKRAWR RR BHSKLWKAGMCCMHWBD RCCWBAUBWRRAKBMRLNMR WHPRCA
KBPFHABA PBCKBCMD WR BBKMCNGH FRSRA CBCWRUMR RLDLRWRAPRR CBW
LCWYHSKAGMCURBAWBPBCYMUWKRCUR WRPRU BHNPRSWCMLRPBH WHAKRAPH
CRAFHSMRWMCNGBNKHPPRALHMKAGMCYMHPBRPBH RY WHAKASMBWKHPCHALH
PRWFY FBYLRWKRALBNMRARCSMLSMHABAAWRYMCWRRDRRAWH CRUBBAMHAWRW
BH BHSKHWBAABCCWCDRBWKHPBR
```

Figure C.13: Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, t variable, Example 2

```
ECATNRAOL TTORCORCT FRUNLC OT TDEG TDNFLC ARGORAOWLG TTORCOR
CT PNGEPERTT TEEP TN YRNG WG LEOUT ORC WNFRCT AT AT OLL TNN
URECALTOWLE AR TDE EORLG TTOYET O UEG GATANRORAET TUREOC TDE
 PETTOYE GATD OR EGORYELALOL KEOL FTERT ORE APURETTEC ORC TE
EV GOGT TN OUULG TDE REG LNRLEUTT TN TDEAR OUULALOTANRT WFT
TDER CATALLFTANRPERT TETT AR TGTTEP LNRGERTANR ORC LFTTNPAKO
TANR OLGOGT URNGET TN WE O WAYYER KNW TDOR ORGNRE ARLLFCARY
PAT LNFLC DOGE URECALTEC LNTTT TNOR GERCNRT URNCFLTT REGER T
EEP TN KFATE LAGE FU TN TDE URNPATET FTERT REGERT TN TDE TRA
EC ORC TRFE ORC CELACE TN LEOGE TDE RATVT TN TDE UANREERT TD
AT TEEPT TN WE TDE LOTE GATD OT LEOTT TDREE NU TDE DAYDER UR
NUALE TTORCORCTPOVARY EUUNRTT ELELTRNRAL COTO ARTERLDORYE EC
A TEE TTNRG U PORFUOLTFRARY OFTNPOTANR URNTNLNLTELDRALOL NUU
ALE URNTNLNL POUTNU TEE TTNRG U ORC NUER TGTTEPT ARTERLNRREL
TANR NTA TEE TTNRG U AR EOLD LOTE TDE PNGEPERT REGER TUREOC
KFATE OT KFALVLG OT UARTT ORTALAUOTEC ECA GDALD URNUNTET TN
PNGE WFTARETT CNLFPEPTT TFLD OT ARGNALET ORC UFRLDOTE NRCERT
 FLELTRNRALOLLG WETGEER LNPUORAET URNGEC TLNGER ORC PNRE EKU
ERTAGE TDOR EKUELTEC EORLG ERTDFTAOTP UNR POU TTORCORCAKEC L
NPPFRALOTANRT UNR CEGALET NR TDE UOLTNRG ULNNR DOT YAGER GOG
 TN TVEUTALATP PNTTLG WELOFTE NU TDE TLNG CEGELNUPERT NU POU
 TTORCORCT ORC URNCFLTT ORC NTA O YLNWOL EUUNRT TN OLDAEGE A
RTERNUEROWALATG WETGEER O GACE RORYE NU LNPUFTER TGTTEPT UOL
ET ROTANROL LNRULALTT WFT AT TDOT ORG REOTNR UNR FTERT TN UR
ET OT TDE TLNG UOLE OLLEUTORLE NU TTORCORCT RNT OT OLL TTORC
ORCT REPOAR O RELETTORG UORT AR TDE EGNLFTANR NU TGTTEPT REY
ORCLETT NU DNG TLNGLG TDE UARTT TTEUT ORE TOVER TDE EKUERTT
DOGE REGATEC TDEAR UNRELOTTT ORC GDALE OLVRNGLECYARY TDOT OL
LEUTORLE AT LNPARY TLNGER TDOR UARTT ORTALAUOTEC AT AT GNRTD
 RNTARY TDOT TDEAR NUTAPATP REPOART TFRGEGT ARCALOTE TNN TDO
T FTERT ARTERETT AR TTORCORCT AT GELL ETTOWLATDEC ATT KFTT T
DOT PNTT NU TDEP DOGE CELACEC TN GOAT UNR EKOPULE GDALE PFLD
 NU TDE GDNLEDEORTEC ERTDFTAOTP AT YNRE ORC TNPE KFETTANREC
TDE GAOWALATG NU POU URNCFLTT TDE FRCERLGARY REEC UNR OR ARC
FTTRG TTORCORC NR TDE UOLTNRG ULNNR REPOART TNPE NU TDE URNW
LEP GALL WE TNLGEC AR TAPE TDRNFYD ECFLOTANR NTA GOTLDERT AR
 TDE UECEROL YNGERRPERT RNTE CAUUERERT LEGELT NU VRNGLECYE O
PNRY CAUUERERT FTERT TDEAR OCGALE AT TDOT UNR RNG FTERT TTOG
 FU TN COTE GATD NTA CEGELNUPERTT LATEROTFRE ORC TN NR NR TD
EG LOR UARC TDEPTELGET KFALVLG UOLLARY WEDARC UNA OLL TDE EK
UELTEC TDNRT TERP URNWLEPT ORC LNTTT OTTNLAOTEC GATD TDE EOR
LG APULEPERTOTANR NU TTORCORCT TDE LNRY TERP YOART TTALL POV
E AT OLL GNRTDGDALE RECFLEC RETEORLD ORC CEGELNUPERT LNTTT U
EGER YOTEGOGT ORC TDEAR OTTNLAOTEC LNTTT ORC RECFLEC UERUNRP
ORLE ORC LNRUNRPORLE TETTARY
```

Figure C.14: Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 3 variables, Example 2

```
ECATNRAUL TTORCORCT FRFNLC OT TDEG TDNFLC ARGORAOVLG TTORCOR
CT PNGEPERTT TEEP TN WRNG VG LEOUT ORC VNFRCT AT AT OLL TNN
URECALTOVLE AR TDE EORLG TTOWET O FEG GATANRORAET TUREOC TDE
 PETTOWE GATD OK EGORWELALOL JEOL FTERT ORE APURETTEC ORC TE
EK GOGT TN OUULG TDE REG LNRLEUTT TN TDEAR OUULALOTANRT VFT
TDER CATALLFTANRPERT TETT AR TGTTEP LNRGERTANR ORC LFTTNPAJO
TANR OLGOGT URNGET TN VE O VAWWER JNV TDOR ORGNRE ARLLFCARW
PAT LNFLC DOGE URECALTEC LNTTT TNOR GERCNRT URNCFLTT REGER T
EEP TN XFATE LAGE FU TN TDE URNPATET FTERT REGERT TN TDE TRA
EC ORC TRFE ORC CELACE TN LEOGE TDE RATKT TN TDE UANREERT TD
AT TEEPT TN VE TDE LOTE GATD OT LEOTT TDREE NF TDE DAWDER UR
NFALE TTORCORCTPOKARW EFFNRTT ELELTRNRAL COTO ARTERLDORWE EC
A TEE TTNRG U PORFFOLTFRARW OFTNPOTANR URNTNLNLTELDRALOL NFF
ALE URNTNLNL POUTNU TEE TTNRG U ORC NUER TGTTEPT ARTERLNRREL
TANR NTA TEE TTNRG U AR EOLD LOTE TDE PNGEPERT REGER TUREOC
XFATE OT XFALKLG OT FARTT ORTALAUOTEC ECA GDALD URNUNTET TN
PNGE VFTARETT CNLFPERTT TFLD OT ARGNALET ORC UFRLDOTE NRCERT
 ELELTRNRALOLLG VETGEER LNPUORAET URNGEC TLNGER ORC PNPE EXU
ERTAGE TDOR EXUELTEC EORLG ERTDFTAOTP FNR POU TTORCORCAJEC L
NPPFRALOTANRT FNR CEGALET NR TDE FOLTNRG FLNNR DOT WAGER GOG
 TN TKEUTALATP PNTTLG VELOFTE NF TDE TLNG CEGELNUPERT NF POU
 TTORCORCT ORC URNCFLTT ORC NTA O WLNVOL EFFNRT TN OLDAEGE A
RTERNUEROVALATG VETGEER O GACE RORWE NF LNPUFTER TGTTEPT FOL
ET ROTANROL LNRFLALTT VFT AT TDOT ORG REOTNR FNR FTERT TN FR
ET OT TDE TLNG UOLE OLLEUTORLE NF TTORCORCT RNT OT OLL TTORC
ORCT REPOAR O RELETTORG UORT AR TDE EGNLFTANR NF TGTTEPT REW
ORCLETT NF DNG TLNCLG TDE FARTT TTEUT ORE TOKER TDE EXUERTT
DOGE REGATEC TDEAR FNRELOTTT ORC GDALE OLKRNGLECWARW TDOT OL
LEUTORLE AT LNPARW TLNGER TDOR FARTT ORTALAUOTEC AT AT GNRTD
 RNTARW TDOT TDEAR NUTAPATP REPOART TFRGEGT ARCALOTE TNN TDO
T FTERT ARTERETT AR TTORCORCT AT GELL ETTOVLATDEC ATT JFTT T
DOT PNTT NF TDEP DOGE CELACEC TN GOAT FNR EXOPULE GDALE PFLD
 NF TDE GDNLEDEORTEC ERTDFTAOTP AT WNRE ORC TNPE XFETTANREC
TDE GAOVALATG NF POU URNCFLTT TDE FRCERLGARW REEC FNR OR ARC
FTTRG TTORCORC NR TDE FOLTNRG FLNNR REPOART TNPE NF TDE URNV
LEP GALL VE TNLGEC AR TAPE TDRNFWD ECFLOTANR NTA GOTLDERT AR
 TDE FECEROL WNGERRPERT RNTE CAFFERERT LEGELT NF KRNGLECWE O
PNRW CAFFERERT FTERT TDEAR OCGALE AT TDOT FNR RNG FTERT TTOG
 FU TN COTE GATD NTA CEGELNUPERTT LATEROTFRE ORC TN NR NR TD
EG LOR FARC TDEPTELGET XFALKLG FOLLARW VEDARC FNR OLL TDE EX
UELTEC TDNRT TERP URNVLEPT ORC LNTTT OTTNLAOTEC GATD TDE EOR
LG APULEPERTOTANR NF TTORCORCT TDE LNRW TERP WOART TTALL POK
E AT OLL GNRTDGDALE RECFLEC RETEORLD ORC CEGELNUPERT LNTTT F
EGER WOTEGOGT ORC TDEAR OTTNLAOTEC LNTTT ORC RECFLEC UERFNRP
ORLE ORC LNRFNRPORLE TETTARW
```

Figure C.15: Cryptanalysis of Monoalphabetic Substitution Using Linear Discriminant, 12 variables, Example 2

```
EDITORIOL STONDORDS UNFOLD OS THEY SHOULD INAORIOWLY STONDOR
DS POAEPENTS SEEP TO BROB WY LEOPS OND WOUNDS IT IS OLL TOO
PREDILTOWLE IN THE EORLY STOBES O FEB AISIONORIES SPREOD THE
 PESSOBE BITH ON EAONBELILOL JEUL USERS ORE IPPRESSED OND SE
EK BOYS TO OPPLY THE NEB LONLEPTS TO THEIR OPPLILOTIONS WUT
THEN DISILLUSIONPENT SETS IN SYSTEP LONAERSION OND LUSTOPIJO
TION OLBOYS PROAES TO WE O WIBBER XOW THON ONYONE INLLUDINB
PIS LOULD HOAE PREDILTED LOSTS SOOR AENDORS PRODULTS NEAER S
EEP TO QUITE LIAE UP TO THE PROPISES USERS REAERT TO THE TRI
ED OND TRUE OND DELIDE TO LEOAE THE RISKS TO THE PIONEERS TH
IS SEEPS TO WE THE LOSE BITH OT LEOST THREE OF THE HIBHER PR
OFILE STONDORDSPOKINB EFFORTS ELELTRONIL DOTO INTERLHONBE ED
I SEE STORY P PONUFOLTURINB OUTOPOTION PROTOLOLTELHNILOL OFF
ILE PROTOLOL POPTOP SEE STORY P OND OPEN SYSTEPS INTERLONNEL
TION OSI SEE STORY P IN EOLH LOSE THE POAEPENT NEAER SPREOD
QUITE OS QUILKLY OS FIRST ONTILIPOTED EDI BHILH PROPOSES TO
POAE WUSINESS DOLUPENTS SULH OS INAOILES OND PURLHOSE ORDERS
 ELELTRONILOLLY WETBEEN LOPPONIES PROAED SLOBER OND PORE EZP
ENSIAE THON EZPELTED EORLY ENTHUSIOSP FOR POP STONDORDIJED L
OPPUNILOTIONS FOR DEAILES ON THE FOLTORY FLOOR HOS BIAEN BOY
 TO SKEPTILISP POSTLY WELOUSE OF THE SLOB DEAELOPPENT OF POP
 STONDORDS OND PRODULTS OND OSI O BLOWOL EFFORT TO OLHIEAE I
NTEROPEROWILITY WETBEEN O BIDE RONBE OF LOPPUTER SYSTEPS FOL
ES NOTIONOL LONFLILTS WUT IS THOT ONY REOSON FOR USERS TO FR
ET OT THE SLOB POLE OLLEPTONLE OF STONDORDS NOT OT OLL STOND
ORDS REPOIN O NELESSORY PORT IN THE EAOLUTION OF SYSTEPS REB
ORDLESS OF HOB SLOBLY THE FIRST STEPS ORE TOKEN THE EZPERTS
HOAE REAISED THEIR FORELOSTS OND BHILE OLKNOBLEDBINB THOT OL
LEPTONLE IS LOPINB SLOBER THON FIRST ONTILIPOTED IT IS BORTH
 NOTINB THOT THEIR OPTIPISP REPOINS SURAEYS INDILOTE TOO THO
T USERS INTEREST IN STONDORDS IS BELL ESTOWLISHED ITS XUST T
HOT POST OF THEP HOAE DELIDED TO BOIT FOR EZOPPLE BHILE PULH
 OF THE BHOLEHEORTED ENTHUSIOSP IS BONE OND SOPE QUESTIONED
THE AIOWILITY OF POP PRODULTS THE UNDERLYINB NEED FOR ON IND
USTRY STONDORD ON THE FOLTORY FLOOR REPOINS SOPE OF THE PROW
LEP BILL WE SOLAED IN TIPE THROUBH EDULOTION OSI BOTLHERS IN
 THE FEDEROL BOAERNPENT NOTE DIFFERENT LEAELS OF KNOBLEDBE O
PONB DIFFERENT USERS THEIR ODAILE IS THOT FOR NOB USERS STOY
 UP TO DOTE BITH OSI DEAELOPPENTS LITEROTURE OND SO ON OR TH
EY LON FIND THEPSELAES QUILKLY FOLLINB WEHIND FOR OLL THE EZ
PELTED SHORT TERP PROWLEPS OND LOSTS OSSOLIOTED BITH THE EOR
LY IPPLEPENTOTION OF STONDORDS THE LONB TERP BOINS STILL POK
E IT OLL BORTHBHILE REDULED RESEORLH OND DEAELOPPENT LOSTS F
EBER BOTEBOYS OND THEIR OSSOLIOTED LOSTS OND REDULED PERFORP
ONLE OND LONFORPONLE TESTINB
```

Figure C.16: Cryptanalysis of Monoalphabetic Substitution by Using Quadratic Discriminant, 1 variable, Example 2

```
ECRTNRROL TTORCORCT FRFNLC OT TDEY TDNFLC RRYORROVLY TTORCOR
CT UNYEUERTT TEEU TN WRNY VY LEOUT ORC VNFRCT RT RT OLL TNN
URECRLTOVLE RR TDE FORLY TTOWET O FEY YRTRNRORRET TUREOC TDE
 UETTOWE YRTD OR EYORWELRLOL QEOL FTERT ORE RUURETTEC ORC TE
EK YOYT TN OUULY TDE REY LNRLEUTT TN TDERR OUULRLOTRNRT VFT
TDER CTTRLLFTRNRUERT TETT RR TYTTEU LNRYERTRNR ORC LFTTNURQO
TRNR OLYOYT URNYET TN VE O VRWWER QNV TDOR ORYNRF RRLLFCRRW
URT LNFLC DOYE URECRLTEC LNTTT TNOR YERCNRT URNCFLTT REYER T
EEU TN XFRTE LRYE FU TN TDE URNURTET FTERT REYERT TN TDE TRR
EC ORC TRFE ORC CELRCE TN LEOYE TDE RRTKT TN TDE URNREERT TD
RT TEEUT TN VE TDE LOTE YRTD OT LEOTT TDREE NF TDE DRWDER UR
NFRLE TTORCORCTUOKRRW EFFNRTT ELELTRNRRL COTO RRTERLDORWE EC
R TEE TTNRY U UORFFOLTFRRRW OFTNUOTRNR URNTNLNLTELDRRLOL NFF
RLE URNTNLNL UOUTNU TEE TTNRY U ORC NUER TYTTEUT RRTERLNRREL
TRNR NTR TEE TTNRY U RR EOLD LOTE TDE UNYEUERT REYER TUREOC
XFRTE OT XFRLKLY OT FRRTT ORTRLRUOTEC ECR YDRLD URNUNTET TN
UNYE VFTRRETT CNLFUERTT TFLD OT RRYNRLET ORC UFRLDOTE NRCERT
 ELELTRNRRLOLLY VETYEER LNUUORRET URNYEC TLNYER ORC UNRE EXU
ERTRYE TDOR EXUELTEC EORLY ERTDFTROTU FNR UOU TTORCORCRQEC L
NUUFRRLOTRNRT FNR CEYRLET NR TDE FOLTNRY FLNNR DOT WRYER YOY
 TN TKEUTRLRTU UNTTLY VELOFTE NF TDE TLNY CEYELNUUERT NF UOU
 TTORCORCT ORC URNCFLTT ORC NTR O WLNVOL EFFNRT TN OLDREYE R
RTERNUEROVRLRTY VETYEER O YRCE RORWE NF LNUUFTER TYTTEUT FOL
ET ROTRNROL LNRFLRLTT VFT RT TDOT ORY REOTNR FNR FTERT TN FR
ET OT TDE TLNY UOLE OLLEUTORLE NF TTORCORCT RNT OT OLL TTORC
ORCT REUORR O RELETTORY UORT RR TDE EYNLFTRNR NF TYTTEUT REW
ORCLETT NF DNY TLNYLY TDE FRRTT TTEUT ORE TOKER TDE EXUERTT
DOYE REYRTEC TDERR FNRELOTTT ORC YDRLE OLKRNYLECWRRW TDOT OL
LEUTORLE RT LNURRW TLNYER TDOR FRRTT ORTRLRUOTEC RT RT YNRTD
 RNTRRW TDOT TDERR NUTRURTU REUORRT TFRYEYT RRCRLOTF TNN TDO
T FTERT RRTERETT RR TTORCORCT RT YELL ETTOVLRTDEC RTT QFTT T
DOT UNTT NF TDEU DOYE CELRCEC TN YORT FNR EXOUULE YDRLE UFLD
 NF TDE YDNLEDEORTEC ERTDFTROTU RT WNRE ORC TNUE XFETTRNREC
TDE YROVRLRTY NF UOU URNCFLTT TDE FRCERLYRRW REEC FNR OR ERC
FTTRY TTORCORC NR TDE FOLTNRY FLNNR REUORRT TNUE NF TDE URNV
LEU YRLL VE TNLYEC RR TRUE TDRNFWD ECFLOTRNR NTR YOTLDERT RR
 TDE FECEROL WNYERRUERT RNTE CRFFERERT LEYELT NF KRNYLECWE O
UNRW CRFFERERT FTERT TDERR OCYRLE RT TDOT FNR RNY FTERT TTOY
 FU TN COTE YRTD NTR CEYELNUUERTT LRTEROTFRE ORC TN NR NR TD
EY LOR FRRC TDEUTELYET XFRLKLY FOLLRRW VEDRRC FNR OLL TDE EX
UELTEC TDNRT TERU URNVLEUT ORC LNTTT OTTNLROTEC YRTD TDE FOR
LY RUULEUERTOTRNR NF TTORCORCT TDE LNRW TERU WORRT TTRLL UOK
E RT OLL YNRTDYDRLE RECFLEC RETEORLD ORC CEYELNUUERT LNTTT F
EYER WOTEYOYT ORC TDERR OTTNLROTEC LNTTT ORC RECFLEC UERFNRU
ORLE ORC LNRFNRUORLE TETTRRW
```

Figure C.17: Cryptanalysis of Monoalphabetic Substitution by Using Quadratic Discriminant, 3 variables, Example 2

ECRTORRAL TTARCARCT FRFOLC AT TLEG TLOFLC RRGARRAVLG TTARCAR
CT UOGEUERTT TEEU TO WROG VG LEAUT ARC VOFRCT RT RT ALL TOO
URECRLTAVLE RR TLE EARLG TTAWET A FEG GRTRORARRET TUREAC TLE
UETTAWE GRTL AR EGARWELRLAL XEAL FTERT ARE RUURETTEC ARC TE
EK GAGT TO AUULG TLE REG LORLEUTT TO TLERR AUULRLATRORT VFT
TLER CRTRLLFTRORUERT TETT RR TGTTEU LORGERTROR ARC LFTTOURXA
TROR ALGAGT UROGET TO VE A VRWWER XOV TLAR ARGORE RRLLFCRRW
URT LOFLC LAGE URECRLTEC LOTTT TOAR GERCORT UROCFLTT REGER T
EEU TO XFRTE LRGF FU TO TLE UROURTET FTERT REGERT TO TLE TRR
EC ARC TRUE ARC CELRCE TO LEAGE TLE RRTKT TO TLE UROREERT TL
RT TEEUT TO VE TLE LATE GRTL AT LEATT TLREE OF TLE LRWLER UR
OFRLE TTARCARCTUAKRRW EFFORTT ELELTRORRL CATA RRTERLLARWE EC
R TEE TTORG U UARFFALTFRRRW AFTOUATROR UROTOLOLTELLRRLAL OFF
RLE UROTOLOL UAUTOU TEE TTORG U ARC OUER TGTTEUT RRTERLORREL
TROR OTR TEE TTORG U RR EALL LATE TLE UOGEUERT REGER TUREAC
XFRTE AT XFRLKLG AT FRRTT ARTRLRUATEC ECR GLRLL UROUOTET TO
UOGE VFTRRETT COLFUERTT TFLL AT RRGORLET ARC UFRLLATE ORCERT
ELELTRORRLALLG VETGEER LOUUARRET UROGEC TLOGER ARC UORE EXU
ERTRGE TLAR EXUELTEC EARLG ERTLFTRATU FOR UAU TTARCARCRXEC L
OUUFRRLATRORT FOR CEGR[illegible]T OR TLE FALTORG FLOOR LAT WRGER GAG
TO TKEUTRLRTU UOTTLG VELAFTE OF TLE TLOG CEGELOUUERT OF UAU
TTARCARCT ARC UROCFLTT ARC OTR A WLOVAL EFFORT TO ALLREGE R
RTEROUERAVRLRTG VETGEER A GRCE RARWE OF LOUUFTER TGTTEUT FAL
ET RATRORAL LORFLRLTT VFT RT TLAT ARG REATOR FOR FTERT TO FR
ET AT TLE TLOG UALE ALLEUTARLE OF TTARCARCT ROT AT ALL TTARC
ARCT REUARR A RELETTARG UART RR TLE EGOLFTROR OF TGTTEUT REW
ARCLETT OF LOG TLOGLG TLE FRRTT TTEUT ARE TAKER TLE EXUERTT
LAGE REGRTEC TLERR FORELATTT ARC GLRLE ALKROGLECWRRW TLAT AL
LEUTARLE RT LOURRW TLOGER TLAR FRRTT ARTRLRUATEC RT RT GORTL
ROTRRW TLAT TLERR OUTRURTU REUARRT TFRGEGT RRCRLATE TCO TLA
T FTERT RRTERETT RR TTARCARCT RT GELL ETTAVLRTLEC RTT XFTT T
LAT UOTT OF TLEU LAGE CELRCEC TO GART FOR EXAUULE GLRLE UFLL
OF TLE GLOLELEARTEC ERTLFTRATU RT WORE ARC TOUE XFETTROREC
TLE GRAVRLRTG OF UAU UROCFLTT TLE [illegible]RCERLGRRW REEC FOR AR RRC
FTTRG TTARCARC OR TLE FALTORG FLOOR REUARRT TOUE OF TLE UROV
LEU GRLL VE TOLGEC RR TRUE TLROFWL ECFLATROR OTR GATLLERT RR
TLE FECERAL WOGERRUERT RCTE CRFFERERT LEGELT OF KROGLECWE A
UORW CRFFERERT FTERT TLERR ACGRLE RT TLAT FOR ROG FTERT TTAG
FU TO CATE GRTL OTR CEGELOUUERTT LRTERATFRE ARC TO OR OR TL
EG LAR FRRC TLEUTELGET XFRLKLG FALLRRW VELRRC FOR ALL TLE EX
UELTEC TLORT TERU UROVLEUT ARC LOTTT ATTOLRATEC GRTL TLE EAR
LG RUULEUERTATROR OF TTARCARCT TLE LORW TERU WARRT TTRLL UAK
E RT ALL GORTLGLRLE RECFLEC RETEARLL ARC CEGELOUUERT LOTTT F
EGER WATEGAGT ARC TLERR ATTOLRATEC LOTTT ARC RECFLEC UERFORU
ARLE ARC LORFORUARLE TETTRRW

Figure C.18: Cryptanalysis of Monoalphabetic Substitution by Using Quadratic Discriminant, 12 variables, Example 2

```
EHITORIAL STANHARHS UNFOLH AS THEY SHOULH INYARIAWLY STANHAR
HS MOYEMENTS SEEM TO GROW WY LEAPS ANH WOUNHS IT IS ALL TOO
PREHICTAWLE IN THE EARLY STAGES A FEW YISIONARIES SPREAH THE
 MESSAGE WITH AN EYANGELICAL XEAL USERS ARE IMPRESSEH ANH SE
EK WAYS TO APPLY THE NEW CONCEPTS TO THEIR APPLICATIONS WUT
THEN HISILLUSIONMENT SETS IN SYSTEM CONYERSION ANH CUSTOMIXA
TION ALWAYS PROYES TO WE A WIGGER XOW THAN ANYONE INCLUHING
MIS COULH HAYE PREHICTEH COSTS SOAR YENHORS PROHUCTS NEYER S
EEM TO QUITE LIYE UP TO THE PROMISES USERS REYERT TO THE TRI
EH ANH TRUE ANH HECIHE TO LEAYE THE RISKS TO THE PIONEERS TH
IS SEEMS TO WE THE CASE WITH AT LEAST THREE OF THE HIGHER PR
OFILE STANHARHSMAKING EFFORTS ELECTRONIC HATA INTERCHANGE EH
I SEE STORY P MANUFACTURING AUTOMATION PROTOCOLTECHNICAL OFF
ICE PROTOCOL MAPTOP SEE STORY P ANH OPEN SYSTEMS INTERCONNEC
TION OSI SEE STORY P IN EACH CASE THE MOYEMENT NEYER SPREAH
QUITE AS QUICKLY AS FIRST ANTICIPATEH EHI WHICH PROPOSES TO
MOYE WUSINESS HOCUMENTS SUCH AS INYOICES ANH PURCHASE ORHERS
 ELECTRONICALLY WETWEEN COMPANIES PROYEH SLOWER ANH MORE EXP
ENSIYE THAN EXPECTEH EARLY ENTHUSI'SM FOR MAP STANHARHIXEH C
OMMUNICATIONS FOR HEYICES ON THE FACTORY FLOOR HAS GIYEN WAY
 TO SKEPTICISM MOSTLY WECAUSE OF THE SLOW HEYELOPMENT OF MAP
 STANHARHS ANH PROHUCTS ANH OSI A GLOWAL EFFORT TO ACHIEYE I
NTEROPERAWILITY WETWEEN A WIHE RANGE OF COMPUTER SYSTEMS FAC
ES NATIONAL CONFLICTS WUT IS THAT ANY REASON FOR USERS TO FR
ET AT THE SLOW PACE ACCEPTANCE OF STANHARHS NOT AT ALL STANH
ARHS REMAIN A NECESSARY PART IN THE EYOLUTION OF SYSTEMS REG
ARHLESS OF HOW SLOWLY THE FIRST STEPS ARE TAKEN THE EXPERTS
HAYE REYISEH THEIR FORECASTS ANH WHILE ACKNOWLEHGING THAT AC
CEPTANCE IS COMING SLOWER THAN FIRST ANTICIPATEH IT IS WORTH
 NOTING THAT THEIR OPTIMISM REMAINS SURYEYS INHICATE TOO THA
T USERS INTEREST IN STANHARHS IS WELL ESTAWLISHEH ITS XUST T
HAT MOST OF THEM HAYE HECIHEH TO WAIT FOR EXAMPLE WHILE MUCH
 OF THE WHOLEHEARTEH ENTHUSIASM IS GONE ANH SOME QUESTIONEH
THE YIAWILITY OF MAP PROHUCTS THE UNHERLYING NEEH FOR AN INH
USTRY STANHARH ON THE FACTORY FLOOR REMAINS SOME OF THE PROW
LEM WILL WE SOLYEH IN TIME THROUGH EHUCATION OSI WATCHERS IN
 THE FEHERAL GOYERNMENT NOTE HIFFERENT LEYELS OF KNOWLEHGE A
MONG HIFFERENT USERS THEIR AHYICE IS THAT FOR NOW USERS STAY
 UP TO HATE WITH OSI HEYELOPMENTS LITERATURE ANH SO ON OR TH
EY CAN FINH THEMSELYES QUICKLY FALLING WEHINH FOR ALL THE EX
PECTEH SHORT TERM PROWLEMS ANH COSTS ASSOCIATEH WITH THE EAR
LY IMPLEMENTATION OF STANHARHS THE LONG TERM GAINS STILL MAK
E IT ALL WORTHWHILE REHUCEH RESEARCH ANH HEYELOPMENT COSTS F
EWER GATEWAYS ANH THEIR ASSOCIATEH COSTS ANH REHUCEH PERFORM
ANCE ANH CONFORMANCE TESTING
```

Figure C.19: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 1 variable, Example 2

```
123123123123123123123123123123123123123123123123123123123123123

ERCETRRTW ADORROSRT COUUCRSOASELSG CCTCCRSRRUOSCOYWG CETCHTR
HASLTUEMSOTC ASEMSETSGSUU GG WETST RORSYTCORC AD AC TWC DSTS
LSSHAUETGCESRRSELS ERCLC ADOYST R USU URACSRRCAST CLSSORSELS
 MSTARGESUADC RO SGTCGEWRCRC GETW MCESC TRE CLFREACERSORR AS
EVSUTCT DS RLFWG DCESOEG CUOCSLTC TU TUEAR TSLLCLTDRTCT GGTS
ELSO RRACCLCTAUOMSOTSTEDT CO CGADEMSLTCGERTAUO RORSLMCETRRVR
EAUO RCGRGASLSUGEC TU YS TSYAAGER VUY DCTC TCGTCE COCWGRCOYS
LAC CUGLR LRGESLSSHAUEER CUTTC AUOSSGECHTRT SCTRGCDT CEGSC C
EER TU VCRTS LCGESGFSETSELS FRSMCTEC MCESC SSGERE DS DCESESC
ERSORR TRGESORR RSLARE DS WETUE DCESCACVASETSELS FCSRSESC TU
RASTESLASETSYESELS CRTESUADC RE WETCE DCSSE UG DCESCAACER FR
SUCCESTTRORRCRCLTVRRA EDGTREASELSLTRSRCL ROTR ACEERLLROYS ER
R CEESTTUCYSL RORCGTUEMRRRA TCETROTCSRSLSUETUSLDECUOAUOLSSUD
RCS FRSTULTW MRLTUL CEESTTUCYSL RORSSFSO CGADEMC ACEERLTCOEU
EAUO UTASTES ADSSC FSRRSETUC UOAS TUE RSGSLECE CEGSC CLSSORS
VMCEESOASVMCLVWG RT DRSCE ROTCLASOTSH SHASULCLLSLSULTCEASETS
LTUE GGACOECT RSCCLECEASTMUC RT COGURCST RORSLMRLLRTESSSRESC
 EWECDCTCRCRCLC YSEGSERSLTRLTCREC FRSGSH CCTGESSORR MUCESEVS
ERCRGS TUORSEVSECDERSETRCYSERDCMCRTCL DSSSLTS ADORROSRRVSH U
SMRGRCLTDRTCT DSSSHEURCST UO DCESGTUETRG DCTUC UOASGAUERSUTC
 TU AVEFDRCCTMSLTCELC YSLTCTESSUSELS AWSGSHEUELULMSOTSSUSLTS
 ADORROSRT RORSLSUHMUEASORR TCR R YWSYRC SGUUCTSETSOCUREUE C
OTSCTSESRYAWRTC YSEGSERSO GRRS SROYS TD CULFCEER ACTTSLASGTU
EASOTDRTCOLSLTCGLCLTC YCE CT DCTD TCG RETCSRSGTR MCESC TU UR
ETSOTSELS AWSGSLTUE RLCSLTROCS TD ADORROSRT CSTSOTSOLW ADORR
OSRT REMRRRSO CECSTARCYSLTRE CO DCESEGUCMDRTC TD ACTTSLASCEA
OSRCECT UG USGSTLUULC TUE DRSCE CEEST RCESETVERSELS EWLEREAS
CTUE REGCTER TUEAR UUCEUOADT RORSULCCESOCVOTGCERGACG DCTD TU
LESETCLESRASLTRRRA AWSGSC DCTC UCCAD TCEAURFREER AD AC GUCTU
 RUEACG DCTD TUEAR TSEARRAR SSLTCOASTMRGECT CORCLTDE DSTSELR
E CTERT COTSCECE CO CETCHTRHASRASUEWC STTRYLCTLSH CEASVMCE D
CTD MUTTSSUSELSL UOGS RSLARERSETSUTCE DSSSEVRLFWE GCAWE RGCU
 TD TUE GCTWELSOSDERSERDCMCRTCL CT ASRS TCH CSMS VCEADRTCERS
ELS GCOYCCADG UG ROFSLSUHMUEASELS MCHERCYCOYSOESH DSSSORSRRR
GADCYSTTRORRCRSSRSELS URLTUCYSGLUSSSCEROACT CSMS TD TUE SCTG
CER GCCLSYESTTWGER AC TCLESELRSMAC SHMUOTCSRSSAC GRECUESC AC
 TUE DERSCTW YUGEROMSOTSOTDE RRUDESSOTSCEUELC TD VCSGWERAE R
LTCG RRUDESSOTSGASCASELSRSSORURCS AC TUOTSGTR RUU CTERT CETC
 MS TU RREESUADC UTASHEUELULMSOTC LCEEROTCCESORR AU TC TR TU
EYSLTC UCORSELSLASCGST UGAUVLC URCLCOYSYEURRR UUC RCLSELS EW
LEUEER AUSSD TSCMSLSUYLSLASORR CUTTC TCTTURTDERSUADC DCESETR
CYSRMSCERERDOTCSRSSUSTTRORRCRC TUE WSRA TSCMSGTCOASTTCCLSLTV
E CE RCLSUTRELGCAWE RERCLER SSTERCCU TCH REGSCTSLECE USADT D
EGSC AOTSUTCT RORSELSRSSOACSCCOTSH USADT RORSCERGCSH SESDSSR
ORUE RORSLTCGTRLTCLESEECEACEA
```

Figure C.20: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 3 variables, Example 2

```
123123123123123123123123123123123123123123123123123123123123123

ETCC WCGRSRCGXUGUUYLUWCSCTRGRKCLRLLDK UCTRCXSGUCGGRLLDCGUHGW
HRRO SEYRWEDSRREYRC RXUSELVLLREGUYLXWTRZ UWTDSACSADSGRCLCL R
YURHARCGVCVRCXRCLRSVXZOVSRCGVRYLXSCRELSCRCLXXZARYLDYURGTRCLR
SYRYRXXVREACKLXWLRTGUXVRCKXCLSEGRSUDEUDSGWELCOAWERDETRGXUSRR
ERREGVYLCLLXYARLLCKVRWVDSKSWKRYEDSESSEVEAWSGUYOC GCC UYLVRER
CLRWLUCRCCOUYASWYRWERYVCYLCWLDLRCEYR  UTVWYASWLXWTR UDC CCKX
CASWLXCXXLRRYUSTVDSESSGRSGRZAWXVWSMSZLCKGUSGUL UELCWKRRTCWVR
OADSKSROUSLXTVRYURHARCVUSKSYEDSRSGURTVUH WYLUZ URKCYLUEFRZLD
EVCSESSTUCERSOCTVRRARC RCLRSAWLYCYVDSUDEUDSURTVWCLCLLCKVRCUC
ETRGXUSEWRVRGXUSTR AUELCLLREGSELCKVRZADKRRC R[illegible]LRSACLXREUDSEV
CRRYVRORRC RZVRCLRSKXYVREACKLXCLREGDCLCKURELSGLCKVRKAWKVWSAW
LCCCVRYEXWTXZTDOGFCXWSVCG WCRREOR EWLXC LUGEX[illegible]AUCVW LXWVRSVU
CLDEVRYESZXRYLCGXUGGRCUWCXWSGUC CGECLXRYUSC RLOCEKVWARGORLCC
CKRSAWLES  RSYXYESYLDEVRYESZXRYLXWTRLARWLDLRCEYDSAUCVW  UWVR
CASWLSYARYVRSRCLUVSARCXREGRKLRGRRSEVELCLFROVUCLUEFRZLDYURGTR
FUCCVRGRRFUC RRLLXYLCCUDCLXWEC AUGERHLRHARELC LRYUSY DERRC R
O SELVRRCWVDYLULKUOVUCRRYURKLXYLCWFSCKRYLXWTRYUW LXYVRLUUEUD
SVREKCZ UCKXCOVSGRCXREXR  CYGUCVDSAWLFRHLDC DEURGXUSYSZVRERU
EXDCFRSEVGXRERUEKCETREGWCXREXCKUDCGDOLCLUROGUSRCGXUGUUCKRHLR
LYCRXC GCC UYLCLURHVSCKRYLSWLCKVRGGRC WLLCC SZLVGRRXASEXREGV
SESSRFEACCKCYYRO DCOVSGR GUYVRLCRCLRSRRLXRHVSEOSYYRWERLCROGU
SRCGXUGUUYLXWTRYUSHURCRRGXUS DCLXSVRLGXCLRGCSZERC RGKVCVSELC
WERZ UEUXZARCEVSGRCXREXRGLDCTRSUXWVRS CSKSOAUCVWSRVYERORRGGR
ERRWGCC UGOR  UGOC EDSGUCLCYLCKGCSGULLWEGDLXRG WSUDEUDSESSCW
EERGERCLRSRRLXRYGRELX KRYEXWKRS CSRCGXUGUUYLULERGERGORSRCGXU
GUUYLWEYXCXRGLUEKRYRXZXRYGWCLCWLCKVREFSCUCC US CSRVYERORRZVW
GUUCVDYLSGLVLXRYOSEOVSEVELCCUDCLDCVUYLXZVRCGFEXRCLRSVAYVWCRR
KGSELWEFCYVUSEVEAWSCSZVRGRCYLXWTRELCCVRGKFW DCVUXAUXLCKGCSGR
 VUCGU VRCRR  CCXWSRRLXRZLCKGUSCCZRCSGUCARCAXCVUSACSADSXSZEV
SXSCAUXLCKGCSEVEAWS UCACCRCSUROGCWRRYUWTVVYLCWTC GCELCL RCLX
CLUYVWYLCWERZVDCLCWLDCGUHGWHRRCRREVRCLRYEXZOCYLRHLCCRR[illegible]UDCLC
KGCSYSYERLCRCLROLVGFRSTR AUETRC REGCCLCLURERXOARELDKARELCRKV
S CSEVELDK RELRGUCETREXCK[illegible]  Y[illegible]WLXRSGUHLDLYRSTUERCC UETR
CLRSFCGGCCACL[illegible]SGLCGARYUSHURCRR[illegible]RSU[illegible]HVWCXCWVRWVRHLCLURGXRCXU
RRCZXRYEXWTX[illegible]RLXRCLRSCX ESZ[illegible]  VCGAUYLDLYRS CSEVELUZ V
CVCSXCCORZVRY RTVUSAUSE[illegible]OVR[illegible]  [illegible]KL[illegible] [illegible]RGECLXRLRCSXXCKVEUDSA[illegible]
SEVELCETRZGRSVSTVWWYRWERW CE[illegible]  [illegible]CEU[illegible]ERCVSEODS CSRULXRETWELX
O UXLUCCCEURWERRRRZRRCLRC[illegible]  [illegible]SA[illegible]SEVGERG WSXSELUYV[illegible]LDCGV
SUUSESSTXCVREACKLSYARHVSEOSY[illegible]UCCVWGEUZVRCXU[illegible]RSS US WSEV
EXR GUSCCWTRCLRORRCFRYLCRARKOVSCXCOCWVRZVVCXUSCSZLXCORCLRSVA
YVRCVUSRVLUCSERZYRYUSZORORRGXUSKSYEDSGDY RCGCETREACKLCKVREGW
CXRCYUCVCEXCGECLXRLCRYEXWTXZTDSEVELRLXWSERZYRXGCWRRYECCOROGF
ELCCLXCORE WCLDKARELWETU VUSURYVXZKVSGUHLUEFRC UOVUCLRLRCYLC
EXRZLWGEREGVYLXWTRCLRCURGRDLKCGERHLRLRCYLXWTRZVURKRHLUEUCLUC
GXRELXWTR  UG WOGU VRCVDCAUCA
```

Figure C.21: Cryptanalysis of Polyalphabetic Substitution Using Linear Discriminant, 20 variables, Example 2

```
123123123123123123123123123123123123123123123123123123123123

EDLTORIIY SLRNORROS LNMULDORSOTHOY ACOLLDOINURRLRFYY ATILCIR
CSOFOUEMONTA SOEMOTOOGRUF GY YEIOS ONDOWOLNDA IL IA IYL LOOO
CROCIUTIGLEOINOTHO EORLL SLRBOS O MOF UISLONORIOS ACRORDOTHO
 MOSSOGEOFILC ON OYILGEYIAOL GEIY WAERA IRE LFPRESAEDORNO SO
EYOFILS LO OCPYY LCEONEG AUNAOCTA TU TUEIR IOCLLLILIOLS GUTO
THON OISLLLLSIUNMONTOSELS LN AYSLEMOLOLVERSIUN ONDOLWATOOIKO
TIUN OLPOYSOCRUYEA TU FO IOWIIGER JUW LCIL ILYOLE LNAYUDLNBO
FIA AUULO HOYEOCROCIUTEO AUSTA SURROYELCORS OROOUALS LEVOR A
EEO TU KLITO LLYEOUPOTOOTHO PROMLSEA WAERA ROYERT LO LCEOTRL
EDORNO TRUEORNO DOLIOE LO YEIUE LCEORIAKSOTOOTHO PLONOERA TU
ISOSEOFSOTOOWEOTHO AOSEOFILC OT YEIAT LCROE UG LCEOCIICER PR
OMLLEOSTONDORDAFIKINI ELGORTSOELOLTRONLL ORTO ILTERLHONBO EO
I AEEOSTURVOC ORNLGIUTWRINI ILTOORTLONOCRUTONOLLEAUNIURLOOML
IAO PROTULOY MOCTUC AEEOSTURVOC ONDOOPON AYSLEMA ILTERLOLNEU
T  N USIOSEO SLORL POINOEIUC URSO TUE OOVOFELT LEVOR ACRORDO
KWLTEORSOKWLLYYY OS LIRAT ONTLLIORTOC OCIOFHLLHOCRUCOAESOTOO
FOUE GUSLNEAS OOALFELTSOSWUC OS LNVUIAOS ONDOCWRLHOSEOOROERA
 EYEALROLIAOLLL FOTPOENOLOOCILIEA PROVOC ALOGERORNO MUREOEWO
ENAIVO TURNOEWOEALEDOEIRLVOENLCWAIIAF LOROFIO SLRNORROIKOC U
OMOUNLLILIOLS LOROCEUIAOS UN LCEOGIUTORY LLOUR URSOGIUENOFIL
 TU SKEPLIALSMOFOATLL FOLILSEOOMOTHO SYOPOCEUELUCMONTOOMOFIO
 SLRNORROS ONDOCRUCWUTSORNO OAI O BYOFOL OGMURTOTOORAUIEUE L
NTOROOEROWIYITL FOTPOENOR GIDO RONBO OL AUFPLTER SLSTOFSOGIU
ESONILIOLRLOLOLGLLLTA FLT LS LCIL ILY REIAONOGOR WAERA TU MR
ETORTOTHO SYOPOCIUE OLAOCTONAO OL SLRNORROS LOTORTORLY SLRNO
RROS REMOINOR LEAOSSORVOCIRT LN LCEOEVULWLIOL OL SLSTOFSOREI
RROLEAS UG UOPOSLUFLL TUE LIRAT ATEOS OREOTIKENOTHO EBCERTSO
CIUE REVLSEO TUEIR MUREURSLS ONDOFHLLEORAKNOGLEOGILG LCIL IU
LEOTILLEOISOLOOINI SYOPOR LCIL MLRSL ILTIUIPOTEO IL IA PURTU
 NUTILG LCIL TUEIR OOTIOISO ROFILNSOSWRYELS LNDLLILE LOOOTHO
T LSERS LNTOREAT LN ATILCIRCSOISOFEYL OSTOWLLSHOC LTSOJWAT L
CIL MUSTOOMOTHOF URVO DOLIOEDOTOOFILT LOROEWOFPYE GCIYE OUAU
 OL TUE GCOYEHORRLEDOENLCWAIIAF LS IONO ILC AOMO KLESLIOLEDO
THO VLRFLLILY UG ORPOCRUCWUTSOTHO WLCERLVLNBONEOC LORORNOINO
USLRVOSTONDORDOONOTHO MOLTURVOGLUOROREORILS AOMO OL TUE OROG
LEO PLLLOWEOSOYYEO IL TLFEOTHROWIC OCWURTLONOOSL POTAUERA IL
 TUE LEDORIY BUYERNMONTONOLE OIMLERONTOLEUELA OL YLOPYEDIE O
FOLG OIMLERONTOUSORSOTHOIRORDUIAO IA TURTOGOR NUF LSERS ATIL
 WO TU DOTEOFILC USIOCEUELUCMONTA LLTERRTLREORNO SU OL OR TU
EVOLIL MLNDOTHOFSOLVOS UUIUKLL MOLLLNBOWEUINO MUR OLLOTHO EB
CEUTEO SUORL TORMOCRUWLOFSORNO AUSTA IASOUIILEDOFILC LCEOEIR
LVOIMOLEOENLRTLONOOMOSTONDORDA TUE YONI TORMOGILNSOSTLLLOFIK
E LT OLLOFORTHGCIYE REDLLEO ROSEORAU ILC OEVOLOOFELT UOSLS L
EPOR IRTOFILS ONDOTHOIRORSAOALRTOC UOSLS ONDOREOUAOC OERLORO
RNJE ONDOLOLGORFILLEOTEATILTI
```

Figure C.22: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 1 variable, Example 2

12312312312312312312312312312312312312312312312312312312312312312312313

ERREARRAV ADORRORRT LOUFLRROAREDRG DLALLRRRRUORROVVG DEALMAR
MARLAUEUROTD AREUREARYRFF BG VEART RORRWALORD RD RD AVL DSAR
LRRMRFEABLERRRREDR ERCDD ADOVRT R URF URARSRRCRRT DLRRORREDR
URTARYERFRDL RO RYALYEVRDRL VEAV UDERD ARE RLFREADERRCRR AR
EKRFADT DS RLFVG DLEROEB DFODRLTD TF TUERR ARLDRLADRALT BGTR
EDRO RRARLDLTRFOUROTRTEDT RO DGADEURLALYERTRFO RORRLUDEARRXR
ERFO RLYRGARLRFYED TF VR ARWRZYER JFW DLAL ALGALE RODVGRROVR
LRD DFGDR DRYERLRRMRFEER DFTTD AFORRYELMART RCARGDDT LEYRC D
EER TF XLRTR DRYERGFREAREDR FRSURTED UDERD RRYERE DS DLERERR
ERRORR TRGERORR RRLRRE DS VEAUE DLERCRDXAREAREDR FRSRRERD TU
RARTERLAREARWEREDR DRTERFRDL RE VEADE DLRRE FG DLERLRZLER FR
SURLERTTRORRCRDLAKRRZ EDGAREAREDRLTRSRRL ROTR RLEERLDROVR ER
R DEERTTFCWRL RORLGAFEURRRZ ALEAROTRSRRLRFEAFSDDEDUORFODRSUD
RDR FRSTFLAV URLTFL DEERTTFCWRL RORRSFRO DGADEUD RLEERLALOEF
ERFO FTRRTER ADSRD FRRRREAFL FOAR TUE RSYRLELE LEYRC DLRRORR
XUREEROARXURLKVG RT DRRDE ROTRLRROTRM RMRRFDRLDRLRFLADEAREAR
LAUF BGAROEDT RSDLLELEARTUFL RT ROYFRDRT RORRLURLDRTERSRRERD
EVEDDCALRDRLDD VREYRERRLARLALRED FRSYRM DLABERRORR UFCEREKR
ERDRYR TUORREKREDDERREARLWRERDLUDRADL DSRRLAR ADORRORRRXRM F
SURGRRLADRALT DSRRMEURDRT FO DLERGAFEARG DLAFC UOARYRUERRFAD
TF AKEFDRDRTURLADEDD VRLALTERSUREDR AVSYRMEUEDFLUROTRSURLAR
ADORRORRT RORRLRFMUFEARORR ADR R VVSVRL RGUFCTREARODUREUE R
OTRCARERRWRVRTD VREYRERRO BRRR RROVR AD DPLFLEER ADTTRLARGAF
EAROADRALODRLALGDRLTD VLE RT DLAD ALG READSRRGAR UDERD TF UR
ETROTREDR AVSYRLAFE RLDRLTRODR AD ADORRORRT LSTROTRODV ADORR
ORRT REURRRRO LEDRTARCWRLARE RO DLEREYFLUDRAL AD ADTTRLARCEZ
ORRLEDT FG USYRTDFFDD TUE DRRDE DEERT RCEREAKERREDR EVLEREAR
LAUE REYRTER TUERR UFCEFOADT RORRFDRLERODKOABLERYRLY DLAD AF
LEREALLERRARLARRRZ AVSYRC DLAL URCAD ALERFRFREER RD RD YFCTU
RFERLY DLAD TUERR ARERRRAR RRLAROARTURYEDT RORRLADE DSAREDR
E LTERT ROTRCEDE RO DEALMARMARRARFEVL RTTRWDRTDRM REARJUDE D
LAD UFTTRSUREDRL UOYR RRLRRERLEARFARE DSRREKRLFVE BLRVE RGDU
AD TUE BLAVEDRORDERREROLUDRADL RT ZSRR ALM DSUR XLEADRALERR
EDR YROVRLRDG FG ROFRLRFMUFEAREDR ULMERLWROVROERM DSRRORRRRR
GADCWRTTRORRCRRSRREDR URLTFCWRGDFSRRCERORLT DSUR AD TUE RCAB
LER YRLDRWERTAVYER RL TRLEREDRSUZL RMUFOTRSRRSAR YREDUERD RL
TUE DERRCAV VFYEROUROTROADE RRUDERROTRLEUEDD AD KLSYVERZE R
LALY RRUDERROTRGARCAREDRRRRORURDR RD TUOTRGAR RFF LTERT DEAD
UR TF RREERFRDL FTRRMEUEDFLUROTD DREEROTLCERORR AF AL AR TU
EWRLAL URORREDRLARLYRT UGRFXDD URLDROVRWEURRR UFC RLDREDR EV
LEFEER AUSRD TRCURLRFWDRLARURR DFTTD ADTAFRADERRFRDL DLEREAR
LWRRURLERERDOTRSRRSURTTRORRCRD TUE VSRZ TRCURYAROARTTRLDRLAK
E RE RLDRFAREDBLRVE RERLLER RRTERCDU ALM REYRLARLELE FSADT D
EYRC ZOTRFADT RORREDRRRROADSDROTRM FSADT RORRCERGDRM RERDSRR
ORFE RORRLALGARLALLEREEDERLER

Figure C.23: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 3 variables, Example 2

```
123123123123123123123123123123123123123123123123123123123123123

ERREARSTV ADORROIRT LOUULRIOAIELIG LLALLRISRUOIROWVG LETLLTR
LAILAUEHIOTL AIEHIEAIYIUU GG VETIT RORIYALORL RD RL TVL DNAI
LIILRUETGLEISRIELI ERCLL ADOWIT R UIU USARNRRCRIT LLIIORIELI
 HITARYEIURDL RO IYTLYEVSLRL BETV ULEIL TRE RLFREALERIORR AI
EKIUTLT DN RLFVG DLEIOEG LUOLILTL TU TUERR TILLRLTDSALT GGTI
ELIO RSARLLLTRUOHIOTITEDT RO LGADEHILALYERTRUO RORILULEARSKR
ERUO RLGRGAILIUYEL TU WI TIYR YER KUY DLTL TLGALE ROLVGRROWI
LRL LUGLR LRYEILIIILRUEER LUTTL AUOIIYELLART ICARGLDT LEGIC L
EER TU KLSTI LRYEIGFIEAIELI FRNHRTEL ULEIL IIYERE DN DLEIEIR
ERIORR TRGEIORR RILRRE DN VETUE DLEICRLKAIEAIELI FRNRIEIL TU
SAITEILAIEAIYEIELI LRTEIURDL RE VETLE DLIIE UG DLEILR LER FR
NURLEITTRORRCRLLTKSR  EDGAREAIELILTRNRRL ROTR RLEERLLROWI ER
S LEEITTUCWIL RORLGTUEURSR  TLEAROTRNRILIUEAUNLDELUORUOLINUD
SLI FRNTULAV HRLTUL LEEITTUCWIL RORINFIO LGADEHL RLEERLALOEU
ERUO UTRITEI ADNIL FISRIETUL UOAI TUE RNGILELE LEGIC LLIIORI
KUREEIOAIKURLKVG RT DSILE ROTRLRIOTIL ILRIULRLLILIULALEAIEAI
LAUE GGAROELT RNLLLELEAITUUL RT ROGUSLIT RORILURLLRTEINIREIL
 EVELDCALSLRLLL WIEGIERILARLTLSEL FRNGIL LLAGEIIORR HUCEIEKI
LRLSGI TUORIEKIELDERIETRLWIERDLULSTLL DNIILTI ALRROIRSKIL U
NHRGRRLTDSALT DNIILEUSLIT UO DLEIGTUEARG DLAUC UOAIYRUERIUTL
 TU AKEFDSLRTHILALELL WILTLTEINUIELI AVNGILEUELULHIOTINUILTI
 ADORROIRT PORILIULUUEAIORR ALS R WVNWRL IGUUCTIEAIOLUSEUE R
OTICAIEIRYRVSTL WIEGIERIO GSRI IROWI AD LULFLEER ALTTILAIGTU
EAIOTDSALOLILALGLRLTL WLE RT DLTD TLG RETLNRIGAR ULEIL TU UR
ETIOTIELI AVNGILTUE RLLILTROLI AD ADORROIRT LNTIOTIOLV ADORR
OIRT RCHRSRIO LELITARCWILTRE RO DLEIEGULUDSAL AD ALTTILAICE
OIRLELT UG UNGITLUULL TUE DSILE LEEIT RCEIETKERIELI EVLEREAI
LTUE REGRTER TUERR UUCEUOADT RORIULRLEIOLKOAGLERYRLY DLTD TU
LEIETLLEISAILARSR  AVNGIC DLTL URCAD TLERUSFREER RD RL GUCTU
 RUERLY DLTD TUERR AIERRSAR IILTROAITURYELT RORRLTDE DNAIELR
E LTERT ROTICELE RO LETLLTRLAISAIUEVL ITTRYLRTLIL REAIKULE D
LTD HUTTINUIELIL UOGI RILRRERIEAIUTRE DNIIEKRLFVE GLRVE RGLU
 AD TUE GLAVELIOIDERIERDLULSTLL RT  NRI TLL LNHI KLEADSALERI
ELI GROWRLRDG UG ROFILIULUUEAIELI ULLERLWROWIOEIL DNIIORISRR
GADCWITTRORRCRINRIELI URLTUCWIGLUNIICERORLT LNHI AD TUE ICAG
LER GRLLIYEITAVYER RL TRLEIELRNU L ILUUOTRNRINAR GRELUEIL RL
 TUE DERICTV WUYEROHIOTIOADE RSUDEIIOTILEUELL AD KLNGVER E R
LALY RSUDEIIOTIGAICAIELISIIORUSLI RL TUOTIGAR RUU LTERT LETL
 UI TU RREEIURDL UTRILEUELULHIOTL LREEROTLCEIORR AU AL AR TU
EWILTL URORIELILAILGIT UGRUKLL URLLROWIYEUSRR UUC RLLIELI EV
LEUEER AUNID TICHILIUYLILAIORR LUTTL TLTAUSTDERIURDL DLEIETR
LWISHILERERDOTRNRINUITTRORRCRL TUE VNR  TICHIYTROAITTRLLILTK
E RE RLLIUARELGLRVE RERLLER IITERCLU TLL REGILAILELE UNADT D
EGIC  OTIUTLT RORIELISIIOALNLROTIL UNADT RORICERGLIL IEIDNIR
ORUE RORILALGARLTLLEIEELERLER
```

Figure C.21: Cryptanalysis of Polyalphabetic Substitution by Using Quadratic Discriminant, 20 variables, Example 2

```
123123123123123123123123123123123123123123123123123123123123

WDOWCOMBYKHUMCRMNRD LNPULDOMHOWHOY UBCLLDOMCUMNOMBYY UWBOGBO
GHOBCUWHONAUKHOWHOWCOGNUB PY YWBOD ONDOBCLNDUKCUKCUKBYL UOCO
CNOGCUWBPLAOMCOWHOKAOLLLKHUMGOD OKPOB UMHOOCOLCOD UCNOMDOWHO
KHODHOGAOBCUB ON OYBOGAYMLOL YWBYKBUWNUKBOW OBUOWHUWDOMCRKHO
WKOBBLD UO OCUYY UBAONAFKLUNLOCAUKAUKAUWCOKBOCLOLBUMCOD PUAO
WHON RMHOLLLDCUNHONAODAUD ON UYHUWHOLCOYAODCUN ONDOLBUWCOMKO
WCUN OLUOYHOCNUYAUKAUKBOKBOBCKGAOKZUB UBBOKBOYCOW ONLYUDONGO
BCUKLUULRKHOYAOCNOGCUWARKLUDAUKHUMNOYAOGCOD OLCRULUD OWVOL U
WAOKAUKXLMAOKLOYAOUUOWCOWHOKUOOHODAUKBUWNUKNOYAOW UO UBAOWNO
WDOMCRKAOUAOMCRKDOLCRW UO YWBUW UBAOLCUKHOWCOWHOKUOOCOWNUKAU
MHODAOBHOWCOBAOWHOKLODAOBCUB OW YWBUW UBNOW UG UBAOBCKBAOKUO
OPOLAODAONDOLDUBBVMCKKAPGCOWHOWLOLAOOCOL RMAOKCOWAOLHONGOKAR
M UWAODAULYOC OMCLGBUWBOMCKKBLWCOMAOOCOCNUWCUOLUWLUNCUMLOOPP
MLOKUOOAULCYKHOCAUC UWAODAULYOC ONDOOUON UYHUWHUKCOWAOLCONAU
WCUN UDCODAOKHUONLKUOMCOWBUB UMHOKAUW OOVOBAOW OWVOL UCNOMDO
KBOWAOMHOKBOLKYY OD PMNUW ONAOLCOMAOG OGCOBHOLHOCNUCCUWHOWCO
BCUW PUHONAUD ROLLBAOWHODBUB OD ONVUMLOD ONDOCBOLHODAOONRWNU
KAYWLULCOMLOLLLKBOWUOWCOLCOCBOMAUKUOOVOG ULCFWNOMCRKHULAOWBO
WCUMVOKAUMCOWBOWLUWDOWBOLYOWCUBBUMBUB PONOBBOKHUMCRMNRMKOG U
OHOUCOLBUMCOD PONOGAIMLOD UN UBAOGBUWCOY PLCUL UMHOGCUWCOBBL
KAUKHVWUUMLODHOBCUWLLKBOLBLDAOOPOWHOKHYOUOGAUWLUCHONAOOPOBBO
KHUMCRMNRD ONDOCNUGBUWHOMCRKCUM OKGYOBOL OGPULAOWCOMLUMAUW O
NAOLCOWNOBCYMALKBOWUOWCOM FMDOKNONGOKCPKLUBULWAOKHLDAOBHOGBU
WHONBUMCOMLOLCOGLOLAUKBLW OD UBBUKBOY OWBUOCOGCOKBUWNUKAUKPO
WAOMAOWHOKHYOUOCBUW OLLOCAONLOKCPKHUMCRMNRD OOAOMAOMLYKHUMCR
MNRD OWHOMCOM OWLODHOLYOCBOW ON UBAOWVULBUMCOKCPKHLDAOBHOLAK
MNRLAUD UG UOUODLUBLLKAUW PMNUW UWAOD OLAOWBVWCOWHOKABCAOWHO
BBUW OWVODARKAUWCOKPULAUMHUD ONDOBHOLAOMLVNCFLARGCOG UBBUKBU
LAOWBOLAOMHOLCOMCKKHYOUOL UBBOKPOLHUKBOWCUMUOWARKCUKCUKUULAU
KCUWCOG UBBUKAUWCOKCOWCOMHOKNOBBONHODBOYALD ONDOLBUW UOCOWHO
W LDAOD ONAOLAUW ON UWBOGBOGHOMHOBAYL ODAOBLODHOG OWHOKBUW U
BBUKHUDAOOPOWHOB UMVOKDOLCRWDOWCOBBOW PONOWBOBUYW FBCYW OULU
KCPKAUW FBCYWHOMNUWDOWCUBBUMBUB OD KOCOKBOG UOHOKXLWHUMCOWDO
WHOKVOMBOLCUY UG OMUOCNUGBUWHOWHOKBOGAOLYONGONAOG PONOMCOMCR
UHULYODAONDOLDOOCOWHOKPOLAULYOGLUONOLAOMCOD UOHOKCPKAUW OLCP
LAOKUOLLOBAODCYYARKCOKAOBAOWHOOBKB OGBUMAOOCOOHOKUOWLUWNUKCO
KAUW PWDOLBYKGUYAONHONAONCUW RMPPWNONAOLAUWLUKCPKKOOUYWDKW O
BCOG RMPPWNONAOUHOLHOWHOMNOMDUMLOKCUKAUMAOGCOKCUB LDAOD UWBL
KBOKAUKDOWAOBCUB UDCOGAUWLUCHONAUKLOWAOMALLAOMCRKHUKCOKCOKAU
WYOLBOKPONDOWHOBHOLVOD UUCUKLLKPOLLONGOBAUMCRKPUL OLLOWHOKAB
CAUWARKHUONUKAOLHOCNUBLOBHOMCRKLUDAUKBUDCUMBUWDOBCUB UBAOWBO
LYOMHOLAOWCUMAOOCOOPODAONDOLDUKAUW YOCKKAOLHOGBONHODAOLLOBBV
W OW OLLOBCOWHFBCYW OWDLLARKNODAOLLUKBOG RWVOLCOBAOW UOHUD P
WUOL KMAOBBLD ONDOWHOMNOMHUOLOMAOG UOHUD ONDOLARULOG OWNPONO
MCUW ONDOLCOGCOBBOLAOWAUWCOWC
```

REFERENCES

[1] A.A. Afifi and S.P. Azen : *Statistical Analysis : A Computer Oriented Approach.* Academic Press, New York. 1979.

[2] T.W. Anderson : *An Introduction to Multivariate Analysis.* John Wiley & Sons, New York. 1958.

[3] C.E. Bosworth : Translation of the Section on Codes and Their Decipherment in Qalqashandi's Subh al-aśhā, *Journal of Semitic Studies, VIII*, 1963.

[4] K.D. Broota : *Experimental Design in Behavioral Research.* John Wiley & Sons, New York. 1989.

[5] J.M. Carroll and S. Martin : The Automated Cryptanalysis of Substitution Ciphers, *Cryptologia, Vol.10* (1986), 193-209.

[6] J.M. Carroll and S. Martin : Using Artificial Intelligence to Break Simple Substitution Ciphers, *The Cryptogram 52(3):3 and 52(4):3*,(1986).

[7] W.J. Dixon, M.B. Brown, L. Engelman, M.A. Hill, R.I. Jennrich : *BMDP Statistical Software Manual, Vol.I.* University of California Press, Berkeley. 1988.

[8] R.O. Duda and P.E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, New York. 1973.

[9] R.A. Fisher : The Use of Multiple Measurements in Taxonomic Problems, *Annals of Eugenics Vol.7*(1936), 179-188

[10] C.C. Foster : Vowel Distribution As A Clue To Vowel Identification, *Cryptologia, Vol.14*(1980), 355-362.

[11] R.E. Frank, W.F. Massy and D.G. Morrison : Bias in Multiple Discriminant Analysis. *Journal of Marketing Research, Vol.2, August*(1965), 250-258.

[12] K. Fukunaga : *Introduction to Statistical Pattern Recognition*. Academic Press, Inc., San Diego. 1972.

[13] H.F. Gaines : *Cryptanalysis*. Dover Publications, New York. 1956.

[14] D.J. Hand : *Discrimination and Classification*. John Wiley & Sons, New York. 1981.

[15] M. James : *Classification Algorithms*. John Wiley & Sons, New York. 1985.

[16] R.I. Jennrich : *Stepwise Discriminant Analysis*. In *Statistical Methods for Digital Computer*, K. Enslein, A. Ralston and H.S. Wilf, eds. John Wiley & Sons, New York. 1977.

[17] R.A. Johnson and D.W. Wichern : *Applied Multivariate Statistical Analysis*. Prentice Hall, New Jersey. 1988.

[18] H. Jürgensen and D.E. Matthews : Some Results on the Information Theoretic Analysis of Cryptosystems. *Advances in Cryptology. Proceedings of Crypto 83* (1984), 303-356.

[19] A.G. Konheim : *Cryptography : A Primer*. John Wiley & Sons, New York. 1981.

[20] P.A. Lachenbruch and M.R. Mickey : Estimation of Error Rates in Discriminant Analysis, *Technometrics 10(1)* (1968), 1-11.

[21] P.A. Lachenbruch : *Discriminant Analysis*. Hafner Press, New York. 1975.

[22] G.J. McLachland : *Discriminant Analysis and Statistical Pattern Recognition*. John Wiley & Sons, New York. 1992.

[23] C.H. Meyer and S.M. Matyas : *Cryptography : A new dimension in computing data security*. John Wiley & Sons, New York. 1982.

[24] D.F. Morrison : *Multivariate Statistical Methods*. McGraw-Hill, New York. 1967.

[25] N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, D.H. Brent : *SPSS - Statistical Package for the Social Sciences (second edition)*. McGraw-Hill, New York. 1975.

[26] C.R. Rao : Inference on Discriminant Function Coefficients , *Essays in Probability and Statistics*, R.C. Bose *et al*, eds. University of North Carolina and Statistical Publishing Society, Chapel Hill. (1970), 587-602.

[27] L.E. Robbins : *The Automated Cryptanalysis of Product Ciphers*. UWO Master Thesis. 1988.

[28] C.E. Shannon : A Mathematical Theory of Communication. *Bell Syst. Tech. J. 27* (1948), 379-423.

[29] L.D. Smith : *Cryptography The Science of Secret Writing*. Dover Publications, New York. 1955.

[30] A.S. Tannenbaum : *Computer Networks*. Prentice Hall, New Jersey. 1981.

[31] R. Von Mises : On the Classification of Observation Data into Distinct Groups, *Ann. Math. Stat. 16* (1945), 68-73.

[32] B.L. Welch : Note on Discriminant Functions, *Biometrika 31* (1939), 218-220.

[33] H.O. Yardley : *The American Black Chamber*. Ballantine Books, New York. 1981.